A review of operational methods of variational and ensemble-variational data assimilation

RMetS
Royal Meteorological Society

Review Article
A review of operational methods of variational and ensemble-variational data assimilation

R. N. Bannister*
Department of Meteorology, National Centre for Earth Observation, University of Reading, UK

*Correspondence to: R. N. Bannister, Department of Meteorology, University of Reading, Earley Gate, Reading RG6 6BB, UK.
E-mail: r.n.bannister@reading.ac.uk

Variational and ensemble methods have been developed separately by various research and development groups and each brings its own benefits to data assimilation. In the last decade or so, various ways have been developed to combine these methods, especially with the aims of improving the background-error covariance matrices and of improving efficiency. The field has become confusing, even to many specialists, and so there is now a need to summarize the methods in order to show how they work, how they are related, what benefits they bring, why they have been developed, how they perform, and what improvements are pending. This article starts with a reminder of basic variational and ensemble techniques and shows how they can be combined to give the emerging ensemble-variational (EnVar) and hybrid methods. A key part of the article includes details of how localization is commonly represented.

There has been a particular push to develop four-dimensional methods that are free of linearized forecast models. This article attempts to provide derivations of the formulations of most popular schemes. These are otherwise scattered throughout the literature or absent. It builds on the nomenclature used to distinguish between methods, and discusses further possible developments to the methods, including the representation of model error.

Key Words: variational data assimilation; ensemble data assimilation; hybrid data assimilation; flow-dependent background-error covariances; localization; model error; nomenclature

Received 15 January 2016; Revised 28 November 2016; Accepted 9 December 2016; Published online in Wiley Online Library

1. Data assimilation and uncertainty

Dealing with uncertainty is at the heart of data assimilation (DA). Forecast models (here those used in numerical weather prediction, NWP) use initial conditions that are imperfect, and are based upon imperfect representations of physical processes. It is well known that a free-running model will accumulate errors until its forecast is no longer useful (Tribbia and Baumhefner, 2004). The only way to restore usefulness is to allow the model to be influenced by observations (Leith, 1993). DA (Daley, 1991; Kalnay, 2002; Rabier, 2005) is the procedure of maintaining the link between evolving models and reality by updating the model with fresh observations. Researchers are working towards formal and robust mathematical methods that work in ways that are consistent with the model, the data, and their degree of uncertainty. DA is related to approaches used in other fields and which go by different names, e.g. state estimation (Wunsch, 2012); optimization (Biegler, 1997); history matching (Emerick, 2012); retrieval production (Rodgers, 2000); inverse modelling (Tarantola, 2005)), and there are additionally many ways of solving a DA problem. Most known DA methods are based on probabilistic theories (most, if not all, exploit Bayes' Theorem (Lorenc, 1986)) and each is made practical by making approximations.

Traditionally there are three Bayesian-based strategies that allow the DA problem to be solved in approximate (hence sub-optimal) ways. These may be categorized as the following.

(i) Variational DA (Var, or its specific implementation 3D-Var or 4D-Var) implements an algorithm to minimize a cost function. In its basic form, Var gives a single, (quasi-) optimal analysis state based on an *a priori* state (the background or forecast), some observations, and the prescribed Gaussian uncertainty statistics for the background and observations. Var requires linearized versions of the observation operators and 4D-Var requires a linearized version of the forecast model.

(ii) Ensemble DA, based on the ensemble Kalman filter (EnKF), offers *a priori* error statistics from an ensemble instead of from a prescribed source, which changes as the system evolves, and does not require use of linearized models.

(iii) Monte-Carlo methods additionally allow the assimilation of information from sources that have non-Gaussian errors.

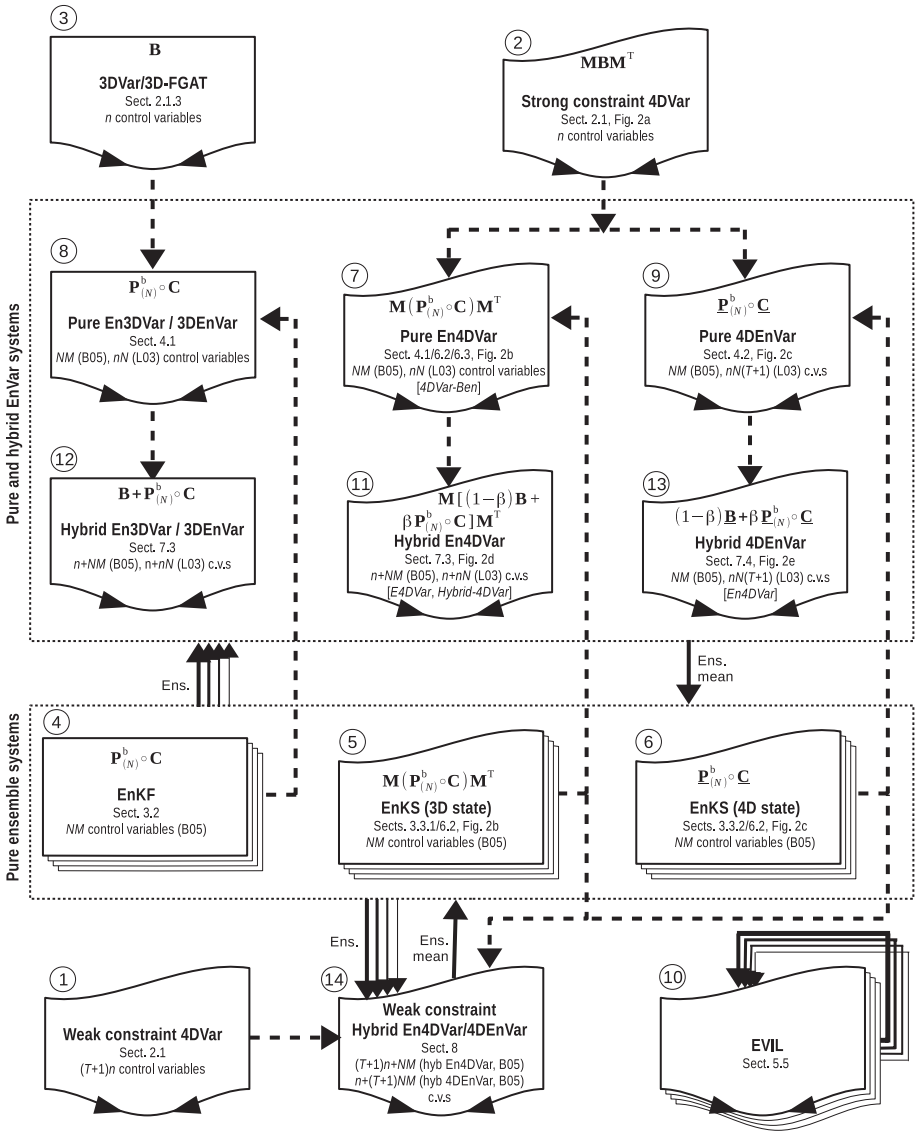

Figure 1. Summary of popular categories of 'pure' and 'combined' DA. A box with a straight (curved) top represents a 3D (4D) scheme (the relevant background-error covariance matrix is specified in the upper portion). A box with a curved (straight) base represents a Var (sequential) scheme. Multiple overlaid boxes are for ensemble schemes. The numbers of control variables are specified and, where localization is used, values are given for the (Buehner, 2005) (B05) and/or (Lorenc, 2003) (L03) methods. The pure Var schemes are boxes ①, ② and ③. The lower dashed box encloses the pure ensemble schemes and the upper dashed box encloses the EnVar schemes. The broad dashed lines show the development path of the theory (for instance hybrid En4DVar (box ⑪) is built from pure En4DVar, ⑦, which is itself built from 4D-Var, ②, and an EnKS scheme, ⑤). The solid arrows show the basic flow of information in a working system, where multiple (single) arrows represent the transfer of an ensemble of states (single state). Alternative names of schemes are in square braces. Acronyms and symbols are defined in the specified sections of this article.

Approaches (i) and (ii) have been successful in NWP, but have limitations (sections 1.2 and 1.3). In recent years this has led to an explosion of techniques that combine them in the hope of maximizing the benefits and eliminating the inadequacies of the separate methods. This includes so-called 'pure EnVar' schemes (section 1.5) and 'hybrid EnVar' schemes (section 1.6). The modern DA toolbox now includes a spectrum of methods that are often subtly different in their name, but may be profoundly different in their treatment of uncertainty (especially in nonlinear systems), in their efficiency, and in the way that the equations are solved (e.g. Buehner *et al.*, 2015b; Liu and Xue, 2016). The boxes in Figure 1 represent the main methods, including Var, ensemble, pure EnVar and hybrid EnVar methods and how they are linked (the Figure caption gives details). The suggested nomenclature to distinguish the methods (Lorenc, 2013) is shown and the Figure will be explained throughout the article.

The aim of this article is to firstly remind ourselves of the basic forms of (i) and (ii), to then review the methods that combine them (pure EnVar and hybrid EnVar), and to summarize the nomenclature. The similarities, differences, advantages, disadvantages and successes of each new method are reviewed. It is beyond the scope of this article to review in depth the numerous ways of implementing each traditional method (such as parallelizable implementations of 4D-Var, and the many implementations of the EnKF such as the stochastic and square-root formulations (Ehrendorfer, 2007; Houtekamer and Zhang, 2016). A review of (iii) is given by van Leeuwen (2009).

The structure of this article is as follows. In the remainder of this section we outline the basic principles of each traditional and combined approach, and mention the advantages, disadvantages and limitations of each. In section 2 we derive common forms of pure variational methods. In section 3 we review the basic principles of pure EnKF-related methods. In section 4 we review how DA systems can be formulated by replacing the fixed background-error statistics of Var with ensemble-derived statistics, while maintaining the variational framework. Broadly, there are two approaches – one that requires the linear models of 4D-Var and another that does not. In section 5 we review hybrid methods, which are DA systems that merge the prescribed and ensemble-derived background-error covariances, again in ways that may or may not require linearized models. For reasons of simplicity, up to this point in this article the inevitable

ensemble sampling noise is not considered. In section 6 we review state-space localization methods which reduce sampling noise. In section 7 we give localized forms of the methods discussed in sections 3–5, and we present a summary of accepted nomenclature. In section 8 we discuss how they could handle model error. In section 9 we compare how they perform in practice, and in section 10 we summarize which schemes are used at leading centres and groups. In section 11 we provide a brief summary and some closing comments. In the appendices we give some important derivations which may help understanding of the theory and efficient implementation of the new assimilation schemes.

1.1. Data assimilation challenges

The objective of DA is to produce information about the posterior probability density function (PDF). This is produced as a first moment (the analysis), or as information about the second moment (e.g. from an ensemble). These can be used as initial conditions for weather forecasts. Observations influence the posterior by matching synthetic versions of the observations (found from the model state via observation operators) to the real observations. This process depends on the nature of each observation (whether it is *in situ* or remotely sensed) and the time of each observation relative to the analysis time. Most observations nowadays are asynoptic (or asynchronous), meaning that respecting the detailed temporal nature of the problem is essential. This problem poses many challenges, including the following.

- Analyses must be produced even in unobserved regions and times, hence the need for prior information (from a background forecast or an ensemble).
- All data are imperfect and uncertainties between different pieces of data can be correlated. The statistics of this information are needed for DA to work well, so uncertain data, for example, should be given lower weight than more precise data. In Var the background- and observational-error covariance matrices are specified and in the EnKF the background-error covariance matrix is estimated from the ensemble. The background-error covariances can significantly affect the analysis (e.g. Navon *et al.*, 2005) and if they are incorrectly specified then the DA is sub-optimal. An incorrect background-error covariance matrix can actually lead to analyses being worse than the prior (Morss and Emanuel, 2002).
- All NWP models are imperfect. This is accounted for in weak-constraint 4D-Var (section 2.1) and in EnKF methods that include stochastic processes in the NWP model (section 3.2).
- Many modes can be excited by DA; these are associated with processes of very different space/time-scales. Modifying the right modes is an issue that goes right back to Richardson's forecast in the 1920s. The use of 4D methods, an appropriate background-error covariance matrix, and initialization (Temperton, 1988) all help with this issue.
- All DA problems for NWP are large and so methods must be able to deal with huge numbers of degrees of freedom (typically $\geq \mathcal{O}(10^7)$ pieces of prior information and similar numbers of observations). Building DA systems that are practicable is often the task that takes the most effort.

The methods dealt with in this review are designed to deal with these points in different ways.

1.2. Variational data assimilation

Variational DA (Talagrand and Courtier, 1987; Schlatter, 2000) is a tool used to estimate a single initial state and a single trajectory of an NWP model.* These are found by minimizing a cost function to optimize the fit of (i) the initial conditions to the background and (ii) the model's version of the observations to the actual observations over a time window.

Variational methods have been used in NWP for many years (Courtier *et al.*, 1994; Park and Zupanski, 2003; Rawlins *et al.*, 2007) as they allow assimilation of a wide range of observations, including those remotely sensed (e.g. from satellite and radar). Variational methods are restricted in their ability to quantify adequately the flow-dependent background and model error statistics, which are prescribed in Var by a parametrized scheme (Trémolet, 2007; Bannister, 2008). Such schemes will almost inevitably be incapable of representing properly the true space and time structure of error statistics. The Var equations are derived in section 2. In 4D-Var the background-error covariance statistics do evolve with the flow within each window, and weak constraint formulations can account for model errors. Normally only the first moment (the mode) of the analysis is found and 4D-Var relies on linearized and adjoint versions of the numerical model.

1.3. Ensemble data assimilation

Instead of specifying a covariance model, the EnKF (Evensen, 1994; Hamill, 2006; Ehrendorfer, 2007; Houtekamer and Zhang, 2016) uses an ensemble of possible forecasts (section 3) that contains valuable flow-dependent information about the background-error statistics. The EnKF is designed to produce an ensemble of analyses, which will similarly contain information about the analysis-error statistics, and can be cycled to future analysis times. The EnKF does not require linearized or adjoint versions of the model, nor of the observation operators. One of the first centres to use an EnKF operationally was Environment Canada (Buehner *et al.*, 2010a,2010b).

Due to cost, the number of ensemble members is too small in practice. Error statistics found from small ensembles will suffer sampling error, and so the statistics will not necessarily be representative of the state. This shows itself in the form of ensemble-derived correlations that are noisy, and variances that are systematically too small (van Leeuwen, 1999; Houtekamer and Mitchell, 2001). The issue with the correlations is problematic when errors between two variables are only weakly correlated in reality (e.g. variables that are widely separated in space).† The issue with the variances will cause the DA to believe that the forecast ensemble is more accurate that it really is. This will place too much trust in the ensemble in comparison with the observations. Left to its own devices, the ensemble will collapse effectively to a single trajectory and the observations will be ignored. This problem in DA is called 'filter divergence' (Houtekamer and Mitchell, 1998). These problems can be lessened with a procedure called localization (which dampens the sample covariances at long range; Gaspari and Cohn, 1999) and ensemble inflation (which artificially increases the spread of an ensemble; Anderson and Anderson, 1999; Hamill *et al.*, 2001) by either multiplicative or additive processes (Kalnay *et al.*, 2007; Whitaker *et al.*, 2008). The *ad hoc* nature of these fixes is largely unsatisfactory,

*Systems that deal with just one realization are often colloquially referred to as 'deterministic' as there is no information about alternative realities (as opposed to systems that deal with ensembles which are referred to as 'probabilistic'). This is a different use of the term 'deterministic' to that used in the theory of dynamical systems. Note that, confusingly, 'deterministic' is also used to describe flavours of the EnKF which do not perturb observations (as opposed to 'stochastic' which is used to describe flavours of the EnKF which do) – section 3.2.

†The expected root-mean-square error in the variance derived from an N-member ensemble is $\sim \sigma^2/N$, where σ^2 is the true variance. The expected root-mean-square error in the correlation between two variables is $\sim (1-\rho^2)/\sqrt{N}$ (Houtekamer and Mitchell, 1998), where ρ is the true value of the correlation. Both errors can be large when N is small and the latter error can be large also when ρ is small.

although methods have been investigated that use adaptive (rather than imposed) moderation functions (Bishop and Hodyss, 2007, 2009a, 2011) and inflation factors that are based on innovation statistics (Bowler *et al.*, 2008).

The EnKF is applied sequentially. The Ensemble Kalman Smoother (EnKS) on the other hand can assimilate observations at times later than the nominal analysis time (Evensen and van Leeuwen, 2000). Elements of the EnKF and the EnKS are covered in section 3.

1.4. Non-Gaussian Monte-Carlo assimilation

Monte-Carlo techniques form a range of methods that include the particle filter (PF). The PF represents a probability distribution by an ensemble (the particles) but, unlike in the EnKF, non-Gaussianity of the distribution is represented. PFs do suffer from degeneracy problems (where the filter ignores all but one particle). For this reason, PFs have to be implemented in special ways which are not used currently for operational purposes (van Leeuwen, 2009). PFs are not considered further in this article.

1.5. Combined 4D-Var/EnKF assimilation schemes ('EnVar')

Most of this article is about combining Var with the EnKF. Such schemes usually run a Var and an EnKF scheme in parallel with information being exchanged between them, which can be one-way or two-way.

One-way coupling can be achieved by shifting each EnKF analysis member for the ensemble mean to equal the Var analysis ('recentring'). In this configuration the EnKF does not transmit information to Var. This kind of coupling has been used, for example, at the Met Office (Bowler *et al.*, 2008).

Another one-way coupling may be achieved by using the ensemble as a source of background-error covariance information (instead of or in addition to climatological covariances) in a Var system to give some flow dependency. This is the basis of En3/4DVar and 3/4DEnVar methods (section 4). Two-way coupling may be achieved, e.g. by combining En3/4DVar or 3/4DEnVar with recentring.

1.6. Hybrid data assimilation

Var is suboptimal partly due to the unrealistic imposed error covariance matrix, and the EnKF is suboptimal partly due to under-sampling. Hybrid DA has been developed to accentuate the best features of each source of background-error information, namely the statistically robust background-error covariances of Var and the flow-dependency of the EnKF. Mathematical approaches have been developed to do this, and most are solved with a (deterministic) Var-style method but with added ensemble information. The simplest solves a Var-like problem, but replaces the static background-error covariance matrix with a weighted average of itself with the error covariance matrix from the ensemble. Efficient methods have been developed to deal with the very large matrices used in real problems, while at the same time attending to the ensemble sampling error problem (section 5).

1.7. Notation

The schemes reviewed in this article try to tackle the above challenges, but they become complex, so a consistent notation is required. We use a notation as close as possible to that of (Ide *et al.*, 1997) and to subsequent literature, although our priority here is to use a common notation throughout the article. To help readers follow the equations, a summary of notation is provided in Table 1.

2. Pure variational data assimilation

Let the n-element vectors $\mathbf{x}(t)$ and $\mathbf{x}^{\mathrm{b}}$ represent a state of the model at time t and the background state at time 0 respectively (quantities with no time argument are implicitly at $t = 0$, and underlined quantities will represent augmented quantities throughout the time window). State $\mathbf{x}(t)$ is found by propagating $\mathbf{x}(t-1)$ forward by one time step δt by the model, $\boldsymbol{\mathcal{M}}_{t-1,t}$:

$$\mathbf{x}(t) = \boldsymbol{\mathcal{M}}_{t-1,t}\{\mathbf{x}(t-1)\} + \boldsymbol{\eta}(t), \tag{1}$$

where the n-element perturbation $\boldsymbol{\eta}(t)$ is to offset model error introduced over $t-1 \rightarrow t$. Let the p_t-element vectors $\mathbf{y}_t^{\mathrm{o}}$ and $\mathbf{y}_t^{\mathbf{x}}$ be the set of observations made at time t, and their modelled counterparts respectively. $\mathbf{y}_t^{\mathbf{x}}$ is related to $\mathbf{x}(t)$ via the observation operator $\boldsymbol{\mathcal{H}}_t$:

$$\mathbf{y}_t^{\mathbf{x}} = \boldsymbol{\mathcal{H}}_t\{\mathbf{x}(t)\}. \tag{2}$$

2.1. 4D-Var cost functions

2.1.1. A full-form (non-incremental) cost function

Var has been the workhorse of DA for many years (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Thepaut and Courtier, 1991; Zupanski, 1997; Rabier *et al.*, 2000; Rawlins *et al.*, 2007). The 4D-Var problem may be posed as a cost function, which is a functional of the $T+1$ states in $\underline{\mathbf{x}} = \{\mathbf{x}, \mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ (over the time window $t = 0$ to T) and then varying $\underline{\mathbf{x}}$ to minimize J (Zupanski, 1997):

$$\begin{aligned} J(\underline{\mathbf{x}}) = &\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{t=0}^{T}\left\|\mathbf{y}_t^{\mathrm{o}} - \boldsymbol{\mathcal{H}}_t(\mathbf{x}(t))\right\|^2_{\mathbf{R}_t^{-1}} \\ &+ \frac{1}{2}\sum_{t=1}^{T}\left\|\mathbf{x}(t) - \boldsymbol{\mathcal{M}}_{t-1,t}(\mathbf{x}(t-1))\right\|^2_{\mathbf{Q}_t^{-1}}, \end{aligned} \tag{3}$$

where $\|\mathbf{a}\|^2_{\mathbf{A}^{-1}} \equiv \mathbf{a}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a}$.

The interpretation of this cost function is as follows.

- The first term (the background term, J_{b}) measures the deviation between $\mathbf{x}$ and $\mathbf{x}^{\mathrm{b}}$. This is calculated in the L^2 norm described by the $n \times n$ background-error covariance matrix $\mathbf{B}_0$.
- The second term (the observation term, J_{o}) measures the deviation between $\mathbf{y}_t^{\mathrm{o}}$ and $\mathbf{y}_t^{\mathbf{x}}$. This is calculated in the L^2 norm described by the $p_t \times p_t$ observation-error covariance matrix $\mathbf{R}_t$ and is summed over the window. Equation (3) does not allow for observations that depend on the model state at more than one time (e.g. precipitation accumulation which at time t depends upon earlier states), although it may be adapted accordingly.
- The third term (the model error term, J_Q) measures the deviation between $\mathbf{x}(t)$ and $\boldsymbol{\mathcal{M}}_{t-1,t}\{\mathbf{x}(t-1)\}$. This is calculated in the L^2 norm described by the $n \times n$ model-error covariance matrices $\mathbf{Q}_t$. In J_Q we have assumed that model errors are uncorrelated in time (white noise approximation) and have *a priori* values of zero.

This minimization may be regarded as an inverse problem, which can be solved given the ability to solve the forward problems ($\boldsymbol{\mathcal{M}}_{t_1,t_2}$ and $\boldsymbol{\mathcal{H}}_t$) and then minimizing $J(\underline{\mathbf{x}})$. Cost function (3) is weak-constraint, meaning that $\underline{\mathbf{x}}$ does not have to be exactly a model trajectory (Sasaki, 1970; Trémolet, 2006). This means that we can formally account for imperfect models in the DA problem.

A common approximation follows by setting $\mathbf{Q}_t = 0$ which forces the states to follow a model trajectory exactly. This is called strong constraint 4D-Var, and such systems have been used operationally for some years (e.g. Rabier *et al.*, 2000; Rabier, 2005; Gauthier *et al.*, 2007; Rawlins *et al.*, 2007). In this approximation only the state at $t = 0$ needs to be determined as all subsequent states follow from the model equations ((1) with $\boldsymbol{\eta} = 0$), and the J_{Q} term no longer appears.

Table 1. Summary of key notation used in this article, where n is the size of the state space, p is the total number of observations over the assimilation window (p_t at time t), T is the number of time steps, N is the number of ensemble members, and $\forall t$ here means 'all times'.

Symbol	Description	Size
$\|\mathbf{a}\|^2_{\mathbf{A}^{-1}}$	$\mathbf{a}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a}$	
$(\delta)\ \mathbf{x}(t)$, $\mathbf{x}$, $\underline{\mathbf{x}}$	State vector: time t, 0, $\forall t$ ($^{\mathrm{b}}$ background, $^{\mathrm{a}}$ analysis, $_{(i)}$ ensemble member)	n, n, $(T+1)n$
$\boldsymbol{\eta}(t)$, $\underline{\boldsymbol{\eta}}$	Model error: time t, $\forall t$	n, Tn
$\boldsymbol{\mathcal{M}}_{t_1,t_2}(\bullet)$	Nonlinear model: $t_1 \rightarrow t_2$	in: n, out: n
$\mathbf{M}_{t_1,t_2}$	Linear model: $t_1 \rightarrow t_2$	$n \times n$
$\underline{\boldsymbol{\mathcal{M}}}(\bullet)$	Nonlinear model: $0 \rightarrow \forall t$	in: n, out: $(T+1)n$
$\underline{\mathbf{M}}$	Linear model: $0 \rightarrow \forall t$	
$\boldsymbol{\mathcal{H}}_t(\bullet)$	Nonlinear observation operator: time t	in: n, out: p_t
$\mathbf{H}_t$	Linear observation operator: time t	$p_t \times n$
$\underline{\boldsymbol{\mathcal{H}}}(\bullet)$	Nonlinear observation operator: $\forall t$	in: $(T+1)n$, out: p
$\underline{\mathbf{H}}$	Linear observation operator: $\forall t$	$p \times (T+1)n$
$\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\bullet)$	Nonlinear observation operator incorporating $\boldsymbol{\mathcal{M}}$	in: n, out: p
$\underline{\mathbf{H}}_{\mathbf{M}}$	Linear observation operator incorporating $\mathbf{M}$	$p \times n$
$\mathbf{y}^{\mathrm{o}}_t$, $\mathbf{y}^{\mathrm{x}}_t$	Observations ($^{\mathrm{o}}$ observed, $^{\mathrm{x}}$ model version, $_{(i)}$ perturbed)	p_t
$\mathbf{B}_0$, $\mathbf{B}^{\mathrm{h}}$	Background-error covariance matrix (time 0, hybrid matrix)	$n \times n$
$\mathbf{R}_t$	Observation-error covariance matrix: time t	$p_t \times p_t$
$\mathbf{Q}_t$	Model-error covariance matrix: time t	$n \times n$
J, J_{b}, J_{o}, J_Q	Cost function, or term therein ($_{\mathrm{b}}$ background, $_{\mathrm{o}}$ observation, $_Q$ model)	
$(\delta)\ \mathbf{d}_t$, $\underline{\mathbf{d}}$	Innovation: time t, $\forall t$	p_t, p
$\mathbf{U}$	Control variable transform for the state	$n \times n$
$\mathbf{V}_t$	Control variable transform for model error: time t	$n \times n$
$\delta\boldsymbol{\chi}_{\mathrm{var}}$, $\boldsymbol{\chi}_{\mathrm{ens}}$, $\boldsymbol{\chi}_{\mathrm{h}}$	Perturbation to the state in control space: time 0 to variational, ensemble, hybrid components	n, N, $n+N$
$\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$, $\widehat{\boldsymbol{\chi}}_{\mathrm{h}}$	Perturbation to the state in control space: time 0 to ensemble, hybrid components	NM, $n+NM$
$\delta\boldsymbol{\gamma}(t)$, $\delta\underline{\boldsymbol{\gamma}}$	Perturbation to the model error in control space: time t, $\forall t$	n, Tn
$\nabla_{\mathbf{v}}$, $\nabla_{\widehat{\mathbf{V}}}$	Gradient vector (with respect to vector $\mathbf{v}$, matrix $\widehat{\mathbf{V}}$)	n, $n \times N$
$\mathbf{X}^{\mathrm{b}}$, $\underline{\mathbf{X}}^{\mathrm{b}}$	Matrix of ensemble members: time t, $\forall t$	$n \times N$, $(T+1)n \times N$
$\widehat{\mathbf{X}}^{\mathrm{b}}$, $\underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$	Matrix of 'localized' ensemble members: time t, $\forall t$	$n \times NM$, $(T+1)n \times NM$
$\mathbf{C}(t)$, $\mathbf{C}$, $\underline{\mathbf{C}}$	Localization matrix: time t, 0, $\forall t$	$n \times n$, $n \times n$, $(T+1)n \times (T+1)n$
$\mathbf{U}^{\mathbf{C}}$	Square root of $\mathbf{C}$ (M is rank of $\mathbf{C}$)	$n \times M$
$\mathbf{P}^{\mathrm{b}}_{(N)}$, $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}$	Sample background-error covariance, localized version	$n \times n$
$\widehat{\mathbf{A}}$, $\widehat{\mathbf{V}}$, $\underline{\widehat{\mathbf{A}}}$, $\underline{\widehat{\mathbf{V}}}$	3D control matrix, precon ditioned version, 4D control matrix, precon ditioned version	$n \times N$, $n \times N$, $(T+1)n \times N$, $(T+1)n \times N$
$\mathbf{Y}^{\mathrm{x}}_t$, $\underline{\mathbf{Y}}^{\mathrm{x}}$	Perturbed observation array at time t, $\forall t$	$p_t \times N$, $p \times N$
$\mathbf{K}^{\mathrm{h}}$, $\mathbf{K}^{\mathrm{ens}}$, $\mathbf{K}^{\mathrm{var}}$	Effective gain matrices for hybrid, EnKF, Var	

In general, an underlined symbol represents that quantity at all relevant times over the assimilation window, and a δ prefix in brackets indicates that the quantity may be represented as a perturbation.

2.1.2. *An incremental cost function*

Cost function (3) is a full-fields approach to DA. A more practical approach is incremental DA, which deals with perturbations made to known reference states (Courtier *et al.*, 1994) and has some advantages over full-fields DA.

The cost function has a quadratic form if $\boldsymbol{\mathcal{M}}_{t_1,t_2}$ and $\boldsymbol{\mathcal{H}}_t$ are linear (which they are not in general). In practical terms, a non-quadratic cost function is difficult to minimize, especially if it possesses multiple local minima. A quadratic problem emerges by linearizing $\boldsymbol{\mathcal{M}}_{t_1,t_2}$ and $\boldsymbol{\mathcal{H}}_t$ about a guess (or reference) state, $\underline{\mathbf{x}}^{\mathrm{g}}$, and formulating the problem in terms of perturbations (increments) to $\underline{\mathbf{x}}^{\mathrm{g}}$. In this article we will take $\underline{\mathbf{x}}^{\mathrm{g}} = \underline{\mathbf{x}}^{\mathrm{b}}$ for simplicity. The result is a modified cost function which is an approximation to the original but is easier to minimize.

The 4D state $\underline{\mathbf{x}}^{\mathrm{b}}$ is found by integrating the model over the window using (1) with the assumption that $\boldsymbol{\eta}(t) = 0$:

$$\mathbf{x}^{\mathrm{b}}(t) = \boldsymbol{\mathcal{M}}_{t-1,t}\left\{\mathbf{x}^{\mathrm{b}}(t-1)\right\}. \tag{4}$$

This can be perturbed to give a general state $\mathbf{x}(t)$ as follows:

$$\mathbf{x}(t) = \mathbf{x}^{\mathrm{b}}(t) + \delta\mathbf{x}(t), \tag{5}$$

$$\boldsymbol{\eta}(t) = \delta\boldsymbol{\eta}(t). \tag{6}$$

Substituting these into (1) gives:

$$\begin{aligned}\mathbf{x}^{\mathrm{b}}(t)+\delta\mathbf{x}(t) &\approx \boldsymbol{\mathcal{M}}_{t-1,t}\left\{\mathbf{x}^{\mathrm{b}}(t-1)\right\} \\ &\quad +\mathbf{M}_{t-1,t}\delta\mathbf{x}(t-1)+\delta\boldsymbol{\eta}(t), \\ \therefore\ \delta\mathbf{x}(t) &\approx \mathbf{M}_{t-1,t}\delta\mathbf{x}(t-1)+\delta\boldsymbol{\eta}(t),\end{aligned} \tag{7}$$

where $\mathbf{M}_{t-1,t} = \partial\mathbf{x}(t)/\partial\mathbf{x}(t-1)$ is the model linearized about $\underline{\mathbf{x}}^{\mathrm{b}}$. Repeated use of (7) leads to:

$$\begin{aligned}\delta\mathbf{x}(t) &\approx \mathbf{M}_{0,t}\delta\mathbf{x}(0)+\mathbf{M}_{1,t}\delta\boldsymbol{\eta}(1)+\mathbf{M}_{2,t}\delta\boldsymbol{\eta}(2)+\ldots \\ &\quad +\mathbf{M}_{t-1,t}\delta\boldsymbol{\eta}(t-1)+\delta\boldsymbol{\eta}(t), \\ &= \mathbf{M}_{0,t}\delta\mathbf{x}(0)+\sum_{\tau=1}^{t}\mathbf{M}_{\tau,t}\delta\boldsymbol{\eta}(\tau),\end{aligned} \tag{8}$$

noting that $\mathbf{M}_{t,t} = \mathbf{I}$. The linearized form of (2) is found in a similar way:

$$\boldsymbol{\mathcal{H}}_t(\mathbf{x}(t)) \approx \boldsymbol{\mathcal{H}}_t\left\{\mathbf{x}^{\mathrm{b}}(t)\right\}+\mathbf{H}_t\left(\mathbf{M}_{0,t}\delta\mathbf{x}(0)+\sum_{\tau=1}^{t}\mathbf{M}_{\tau,t}\delta\boldsymbol{\eta}(\tau)\right), \tag{9}$$

where $\mathbf{H}_t = \partial \mathbf{y}^{\mathbf{x}}_t / \partial \mathbf{x}(t)$. Substituting (9) into (3) gives the incremental cost function:

$$J_{\rm inc}(\delta\mathbf{x}, \delta\underline{\boldsymbol{\eta}}) = \frac{1}{2}\left\|\delta\mathbf{x}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{t=0}^{T}\left\|\delta\mathbf{d}_t\right\|^2_{\mathbf{R}_t^{-1}} + \frac{1}{2}\sum_{t=1}^{T}\left\|\delta\boldsymbol{\eta}(t)\right\|^2_{\mathbf{Q}_t^{-1}}, \tag{10}$$

$$\text{where } \delta\mathbf{d}_t = \mathbf{d}_t - \mathbf{H}_t\left(\mathbf{M}_{0,t}\delta\mathbf{x} + \sum_{\tau=1}^{t}\mathbf{M}_{\tau,t}\delta\boldsymbol{\eta}(\tau)\right), \tag{11}$$

$$\text{and } \mathbf{d}_t = \mathbf{y}^{\rm o}_t - \boldsymbol{\mathcal{H}}_t\left\{\mathbf{x}^{\rm b}(t)\right\}. \tag{12}$$

These equations represent incremental weak-constraint 4D-Var (① in Figure 1). Minimizing $J_{\rm inc}$ (over an inner loop) gives particular values of $(\delta\mathbf{x}, \delta\underline{\boldsymbol{\eta}}) = \{\delta\mathbf{x}, \delta\boldsymbol{\eta}(1), \ldots, \delta\boldsymbol{\eta}(T)\}$. The strong-constraint 4D-Var version is obtained by setting all $\delta\boldsymbol{\eta}(t) = 0$ in (10) and (11), which is valid when model error is negligible over the window (② in Figure 1).

For practical reasons Var uses a $\mathbf{B}_0$-matrix (and weak-constraint 4D-Var uses $\mathbf{Q}_t$ matrices) that is (are) essentially the same for each cycle, although flow-dependence can be added artificially to the way that $\mathbf{B}_0$ is modelled by, for example, cycling variance information (Dee, 2002) and the use of (linearized) nonlinear balance relationships (Fisher, 2003; Barker *et al.*, 2004).

2.1.3. *Further simplifications*

Strong-constraint 4D-Var may be further simplified to exclude part of the time dimension (Courtier *et al.*, 1998; Gauthier *et al.*, 1999; Lorenc *et al.*, 2000; Gustafsson *et al.*, 2001). The most popular is called 3D-FGAT (3D-Var with the First Guess at the Appropriate Time, but also known as just 3D-Var), which emerges from the strong-constraint version of (10) with the assumption that $\mathbf{M}_{t_1,t_2} \approx \mathbf{I}\ \forall t_1, t_2$ (③ in Figure 1):

$$J^{\rm 3DFGAT}_{\rm inc}(\delta\mathbf{x}) = \frac{1}{2}\left\|\delta\mathbf{x}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{t=0}^{T}\left\|\delta\mathbf{d}_t\right\|^2_{\mathbf{R}_t^{-1}},$$
$$\delta\mathbf{d}_t = \mathbf{d}_t - \mathbf{H}_t\delta\mathbf{x}. \tag{13}$$

(Lee *et al.*, 2004; Lawless, 2010), where $\delta\mathbf{x}$ is here valid at the centre of the observation window. The 3D aspect is due to the loss of the time dimension through the neglect of all $\mathbf{M}_{t_1,t_2}$ and the FGAT aspect is due to the preservation of time in $\delta\mathbf{d}_t$ ((12) still holds for $\mathbf{d}_t$). Notably, 3D-FGAT was used to produce the ERA-40 reanalysis (Uppala *et al.*, 2005).

Note that even operational 4D-Var systems in practice use approximations, e.g. the Jacobian $\mathbf{M}_{0,t}$ does not normally contain linearizations of the complete model, as many physical processes are problematic to linearize (Xu, 1996).

2.2. *The derivative of the 4D-Var cost function*

The minimum of $J_{\rm inc}$ is found iteratively (e.g. Liu and Nocedal, 1989) with algorithms that require calculation of the gradient of (10), $\nabla J_{\rm inc}$. This is a column vector of the same structure as the argument of $J_{\rm inc}$ in (10), i.e. $(\delta\mathbf{x}, \delta\underline{\boldsymbol{\eta}})$, but of derivatives as follows:

$$\nabla J_{\rm inc} = \begin{pmatrix} \nabla_{\delta\mathbf{x}} J_{\rm inc} \\ \nabla_{\delta\boldsymbol{\eta}(1)} J_{\rm inc} \\ \vdots \\ \nabla_{\delta\boldsymbol{\eta}(T)} J_{\rm inc} \end{pmatrix}, \tag{14}$$

where the ith components $[\bullet]_i$ of the sub-vectors are:

$$[\nabla_{\delta\mathbf{x}} J_{\rm inc}]_i = \partial J_{\rm inc} / \partial [\delta\mathbf{x}]_i,$$
$$\text{and } [\nabla_{\delta\boldsymbol{\eta}(t)} J_{\rm inc}]_i = \partial J_{\rm inc} / \partial [\delta\boldsymbol{\eta}(t)]_i.$$

There are n components of each sub-vector, so $\nabla J_{\rm inc}$ has $(T+1)n$ components in total. (For strong-constraint 4D-Var and 3D-Var, elements exist with respect to only $\delta\mathbf{x}$ in (14) so $\nabla J_{\rm inc}$ has just n components.) The gradient is important because in phase space $-\nabla J_{\rm inc}$ points in the direction of greatest descent of $J_{\rm inc}$.

Components of $\nabla J_{\rm inc}$ are derived in Appendix A to be:

$$\nabla_{\delta\mathbf{x}} J_{\rm inc} = \mathbf{B}_0^{-1}\delta\mathbf{x} - \sum_{t=0}^{T}\mathbf{M}^{\rm T}_{0,t}\mathbf{H}^{\rm T}_t\mathbf{R}_t^{-1}\delta\mathbf{d}_t, \tag{15}$$

$$\nabla_{\delta\boldsymbol{\eta}(t')} J_{\rm inc} = -\sum_{t=t'}^{T}\mathbf{M}^{\rm T}_{t',t}\mathbf{H}^{\rm T}_t\mathbf{R}_t^{-1}\delta\mathbf{d}_t + \mathbf{Q}_{t'}^{-1}\delta\boldsymbol{\eta}(t'). \tag{16}$$

4D-Var is limited in its scope for computational parallelizability. The estimation of $\delta\mathbf{x}$ and $\delta\boldsymbol{\eta}(t)$ –as in formulation (10) – cannot be parallelized temporally (e.g. for $\delta\boldsymbol{\eta}(t)$: this requires integration of the forward linear model from earlier times in (11), and the integration of the adjoint model from future times in (16)). There are alternative formulations though that may allow parallelization of 4D-Var in time (M. Fisher, 2011; personal communication; Desroziers and Berre, 2012) .

2.3. *Control variable transforms in Var*

The cost function (10), with definitions (11) and (12), requires the matrices $\mathbf{B}_0$, $\mathbf{R}_t$, and $\mathbf{Q}_t$ to be known explicitly, but even modern computers are incapable of dealing with such large matrices. The method of control variable transforms (CVTs) is a 'trick' to represent $\mathbf{B}_0$ and $\mathbf{Q}_t$ without needing to know them explicitly.[‡] Instead of dealing with (10) – a function of $(\delta\mathbf{x}, \delta\underline{\boldsymbol{\eta}})$ –an alternative cost function is formulated in terms of new variables ('control variables'). There are number of key studies that describe the choice of control variables used for meteorological DA (e.g. Parrish and Derber, 1992; Derber and Bouttier, 1999; Berre, 2000), for ocean DA (Weaver *et al.*, 2005), and in general (Bannister, 2008; Ménétrier and Auligné, 2015).

The following transformations may be made for $\delta\mathbf{x}$, and $\delta\boldsymbol{\eta}(t)$ in (10):

$$\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}_{\rm var}, \tag{17}$$

$$\delta\boldsymbol{\eta}(t) = \mathbf{V}_t\delta\boldsymbol{\gamma}_{\rm var}(t). \tag{18}$$

Here $\delta\boldsymbol{\chi}_{\rm var}$ and $\delta\boldsymbol{\gamma}_{\rm var}(t)$ are 'control variables', which are associated with the model space variables $\delta\mathbf{x}$ and $\delta\boldsymbol{\eta}(t)$ respectively, and $\mathbf{U}$ and $\mathbf{V}_t$ are the CVTs. Vectors $\delta\boldsymbol{\chi}_{\rm var}$ and $\delta\boldsymbol{\gamma}_{\rm var}(t)$ need not have n elements each, but for the purposes of this article we shall assume that each is an n-element vector. Substituting (17) and (18) into (10) gives the *preconditioned* cost function:

$$J_{\rm inc}(\delta\boldsymbol{\chi}_{\rm var}, \delta\underline{\boldsymbol{\gamma}}_{\rm var}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\rm var}\right\|^2_{\mathbf{U}^{\rm T}\mathbf{B}_0^{-1}\mathbf{U}} + \frac{1}{2}\sum_{t=0}^{T}\left\|\delta\mathbf{d}_t\right\|^2_{\mathbf{R}_t^{-1}} + \frac{1}{2}\sum_{t=1}^{T}\left\|\delta\boldsymbol{\gamma}_{\rm var}(t)\right\|^2_{\mathbf{V}_t^{\rm T}\mathbf{Q}_t^{-1}\mathbf{V}_t}, \tag{19}$$

and substituting into (11) gives:

$$\delta\mathbf{d}_t = \mathbf{d}_t - \mathbf{H}_t\left(\mathbf{M}_{0,t}\mathbf{U}\delta\boldsymbol{\chi}_{\rm var} + \sum_{\tau=1}^{t}\mathbf{M}_{\tau,t}\mathbf{V}_\tau\delta\boldsymbol{\gamma}_{\rm var}(\tau)\right), \tag{20}$$

[‡]The $\mathbf{R}_t$ matrices are often assumed diagonal. Over all time steps, there are $p = \sum_{t=0}^{T} p_t$ observations in total, which is a practical number of diagonal elements. There is evidence though that the correlations between errors (including representivity) of different observations should be accounted for (Stewart *et al.*, 2014; Waller *et al.*, 2014).

($\mathbf{d}_t$ is still defined by (12)). The aim is to have control variables that have error covariances $\mathbf{I}$, which is achieved by choosing $\mathbf{U}$ and $\mathbf{V}_t$ such that $\mathbf{U}^{\mathrm{T}}\mathbf{B}_0^{-1}\mathbf{U}=\mathbf{I}$ and $\mathbf{V}_t^{\mathrm{T}}\mathbf{Q}_t^{-1}\mathbf{V}_t=\mathbf{I}$. Setting these in (19) gives:

$$J_{\mathrm{inc}}(\delta\boldsymbol{\chi}_{\mathrm{var}},\delta\boldsymbol{\chi}_{\mathrm{var}}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\sum_{t=0}^{T}\|\delta\mathbf{d}_t\|^2_{\mathbf{R}_t^{-1}} + \frac{1}{2}\sum_{t=1}^{T}\left\|\delta\boldsymbol{\gamma}_{\mathrm{var}}(t)\right\|^2_{\mathbf{I}}. \quad (21)$$

The gradients in control space follow from (15) and (16) with the chain rule:

$$\nabla_{\delta\boldsymbol{\chi}_{\mathrm{var}}}J_{\mathrm{inc}} = \mathbf{U}^{\mathrm{T}}\nabla_{\delta\mathbf{x}}J_{\mathrm{inc}}, \quad (22)$$

$$\text{and } \nabla_{\delta\boldsymbol{\gamma}_{\mathrm{var}}(t')}J_{\mathrm{inc}} = \mathbf{V}_{t'}^{\mathrm{T}}\nabla_{\delta\boldsymbol{\eta}(t')}J_{\mathrm{inc}}. \quad (23)$$

Cost function (21) is minimized with respect to $(\delta\boldsymbol{\chi}_{\mathrm{var}},\delta\boldsymbol{\chi}_{\mathrm{var}})$, which is an easier and better conditioned problem than minimizing (10) with respect to $(\delta\mathbf{x},\delta\boldsymbol{\eta})$ (Lorenc *et al.*, 2000; Haben *et al.*, 2011) and useful physical constraints can be built into $\mathbf{U}$ (Ingleby, 2001; Bannister, 2008). Importantly the two problems are equivalent given:

$$\mathbf{B}_0 = \mathbf{U}\mathbf{U}^{\mathrm{T}}, \quad (24)$$

$$\text{and } \mathbf{Q}_t = \mathbf{V}_t\mathbf{V}_t^{\mathrm{T}}, \quad (25)$$

which follow from $\mathbf{U}^{\mathrm{T}}\mathbf{B}_0^{-1}\mathbf{U}=\mathbf{I}$ and $\mathbf{V}_t^{\mathrm{T}}\mathbf{Q}_t^{-1}\mathbf{V}_t=\mathbf{I}$ above. The optimum values of $(\delta\boldsymbol{\chi}_{\mathrm{var}},\delta\boldsymbol{\chi}_{\mathrm{var}})$ relate to $(\delta\mathbf{x},\delta\boldsymbol{\eta})$ with the CVTs (17) and (18). In practice $\mathbf{U}$ and $\mathbf{V}_t$ are modelled (leading to *implied* covariance matrices above) by making simplifying assumptions about how errors are thought to be related (Bannister, 2008), and the CVT viewpoint will be key to understanding EnVar methods (section 4 onwards). In most of the remainder of this article we do not consider model error.

3. Pure ensemble data assimilation systems

Var has been a success in NWP but is limited by the $\mathbf{B}_0$-matrix, which influences the quality of the analysis (Berre *et al.*, 2006). Despite its importance, $\mathbf{B}_0$ has to be estimated – usually crudely – due to its large size. It is also assumed to be static (at the start of each cycle), even though in reality background-error statistics change in time. Ensemble methods (e.g. the EnKF) circumvent this problem by representing background uncertainty with an ensemble. We show briefly how the EnKF approximates background-error statistics and mention some basic EnKF formulations. However, to avoid complications here, we leave discussion of localization until section 6.

3.1. Background uncertainty represented by an ensemble

Consider a population of N background forecasts,[§] each labelled with index (i) $\mathbf{x}^{\mathrm{b}}_{(i)}$ (all valid at $t=0$). The $n\times n$ background-error covariance matrix $\mathbf{P}^{\mathrm{b}}$ is approximated by this population:

$$\mathbf{P}^{\mathrm{b}} \approx \mathbf{P}^{\mathrm{b}}_{(N)} = \frac{1}{N-1}\sum_{i=1}^{N}\delta\mathbf{x}^{\mathrm{b}}_{(i)}\delta\mathbf{x}^{\mathrm{b}\mathrm{T}}_{(i)} = \mathbf{X}^{\mathrm{b}}\mathbf{X}^{\mathrm{b}\mathrm{T}}, \quad (26)$$

$$\text{where } \delta\mathbf{x}^{\mathrm{b}}_{(i)} = \mathbf{x}^{\mathrm{b}}_{(i)} - \overline{\mathbf{x}^{\mathrm{b}}}. \quad (27)$$

Here $\mathbf{P}^{\mathrm{b}}$ is the true (and full-rank) matrix, approximated by $\mathbf{B}_0$ in Var and by $\mathbf{P}^{\mathrm{b}}_{(N)}$ in the EnKF. The overbar in (27) indicates the sample average. $\mathbf{P}^{\mathrm{b}}_{(N)}$ is approximate (rank-$N-1$) due to sampling error as N is finite (section 6). A compact way of writing $\mathbf{P}^{\mathrm{b}}_{(N)}$ is given in (26) where the columns of the $n\times N$ matrix $\mathbf{X}^{\mathrm{b}}$ are $\delta\mathbf{x}^{\mathrm{b}}_{(i)}/\sqrt{N-1}$ (Evensen, 1994).

3.2. Ensemble filters

Substituting $\mathbf{P}^{\mathrm{b}}_{(N)}$ into the Kalman filter (KF) update equation (Jazwinski, 1970; Evensen, 2003) gives:

$$\mathbf{x}^{\mathrm{a}}_{(i)} = \mathbf{x}^{\mathrm{b}}_{(i)} + \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R}+\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{d}_{(i)}, \approx \mathbf{x}^{\mathrm{b}}_{(i)} + \mathbf{X}^{\mathrm{b}}(\mathbf{H}\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}\left\{\mathbf{R}+\mathbf{H}\mathbf{X}^{\mathrm{b}}(\mathbf{H}\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}\right\}^{-1}\mathbf{d}_{(i)}, \quad (28)$$

where $\mathbf{d}_{(i)} = \mathbf{y}^{\mathrm{o}}_{(i)} - \boldsymbol{\mathcal{H}}(\mathbf{x}^{\mathrm{b}}_{(i)})$. Equation (28) is evaluated for each ensemble member ($i=1,\ldots N$), each with a perturbed set of observations, $\mathbf{y}^{\mathrm{o}}_{(i)} = \mathbf{y}^{\mathrm{o}} + \boldsymbol{\epsilon}^{\mathrm{o}}_{(i)}$, where $\boldsymbol{\epsilon}^{\mathrm{o}}_{(i)} \sim N(\mathbf{0},\mathbf{R})$. The ensemble of analyses $\mathbf{x}^{\mathrm{a}}_{(i)}$ can be cycled to the next observation time. The key idea is that (28) does not need the explicit $n\times n$ matrix $\mathbf{P}^{\mathrm{b}}_{(N)}$. Equation (28) is a stochastic formulation of the EnKF (Houtekamer and Mitchell, 1998; Burgers *et al.*, 1998) – so-called because the observations are perturbed stochastically to give the correct spread of the analysis ensemble. Other important variations of the EnKF (e.g. the ensemble transform Kalman filter) use deterministic formulations (Anderson, 2001; Bishop *et al.*, 2001; Whitaker and Hamill, 2002; Tippett *et al.*, 2003) which do not require perturbed observations. (It is outside the scope of this article to review the many flavours of ensemble filters.) The EnKF is ④ in Figure 1.

The last line of (28) is written so that it does not require the adjoint of the observation operator. The linear operation $\mathbf{H}\mathbf{X}^{\mathrm{b}}$ is also often approximated by the nonlinear operations $\left\{\boldsymbol{\mathcal{H}}(\mathbf{x}^{\mathrm{b}}_{(i)}) - \boldsymbol{\mathcal{H}}(\overline{\mathbf{x}^{\mathrm{b}}})\right\}/\sqrt{N-1}$ (approximating the ith column of $\mathbf{H}\mathbf{X}^{\mathrm{b}}$). The EnKF can therefore be implemented without the need for linearized operators (this includes the linearized forecast model in the case of the ensemble smoother – section 3.3). This represents a significant reduction of development effort compared to Var (Lorenc, 2003). Many studies show that the EnKF is a useful tool in NWP, often comparable or better than Var (e.g. Houtekamer *et al.*, 2005; Zhang *et al.*, 2006; Meng and Zhang, 2007; Bonavita *et al.*, 2008; Meng and Zhang, 2008a,2008b; Szunyogh *et al.*, 2008, Whitaker *et al.*, 2008; Hamill *et al.*, 2011a,2011b; Zhang *et al.*, 2011, Houtekamer *et al.*, 2014). The differences between EnKF and Var can be most evident in data-sparse regions (Whitaker *et al.*, 2008).

The background-error covariance matrix in the ordinary KF is assumed to include predictability errors (propagation of analysis error from the previous analysis, generically $\mathbf{M}\mathbf{P}^{\mathrm{a}}\mathbf{M}^{\mathrm{T}}$) and a model error contribution ($\mathbf{Q}$). Since the model producing the $\mathbf{x}^{\mathrm{b}}_{(i)}$ is imperfect, model error should be allowed to affect the spread of the ensemble background states (the system simulation approach of Houkekamer *et al.*, 1996). There are many strategies for accounting for model error, including covariance inflation (Whitaker *et al.*, 2002) and additive error (Houtekamer *et al.*, 2005). These approaches have been tested in an EnKF with a simplified primitive-equation model to represent the model error introduced due to unresolved scales (Hamill and Whitaker, 2005). However the difficulty is choosing the degree of inflation and the structure of the additive error. Hamill and Whitaker (2005) chose to prescribe inflation factors a little over unity (although more advanced approaches are possible, e.g. Raynaud *et al.*, 2012), and they chose additive errors from three options: random samples of differences between forecasts of different resolutions, random samples of differences between forecasts and climatology, and random samples of forecast tendencies. It is also possible to relax the analysis ensemble spread closer to the spread of the forecast ensemble (Zhang *et al.*, 2004).

[§]For operational DA, $N \ll n$, and presently N is $\mathcal{O}(10^1)$ to $\mathcal{O}(10^3)$.

3.3. Ensemble smoothers

In the EnKF, $\mathbf{x}^{\rm b}_{(i)}$, $\mathbf{y}^{\rm o}_{(i)}$, and $\mathbf{x}^{\rm a}_{(i)}$ are valid at $t=0$, but information is propagated from one time to the next via cycling (sequential DA). An EnKS (Evensen, 2007) on the other hand introduces the time dimension into the analysis problem itself so it can deal directly with observations spanning a time window ($0 \le t \le T$). We discuss two ways of implementing an EnKS, which later influence how EnVar is developed.

3.3.1. An EnKS with a 3D state vector

The EnKF (28) may be adapted by keeping its 3D state vector (valid at the start of the window), but incorporating the forecast model into the observation operator (as in 4D-Var) for observations made at $t > 0$:

$$\mathbf{x}^{\rm a}_{(i)} = \mathbf{x}^{\rm b}_{(i)} + \mathbf{X}^{\rm b}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}\right)^{\rm T}\left\{\underline{\mathbf{R}} + \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}\right)^{\rm T}\right\}^{-1}\underline{\mathbf{d}}_{(i)}, \tag{29}$$

where $\underline{\mathbf{d}}_{(i)} = \underline{\mathbf{y}}^{\rm o}_{(i)} - \underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b}_{(i)})$ (see below). Objects (not) underlined are for ($t=0$) $t \ge 0$: $\underline{\mathbf{y}}^{\rm o}_{(i)}$ contains the perturbed observations (error covariance $\underline{\mathbf{R}}$) over the window in chronological order, $\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b}_{(i)})$ is the time-distributed observation operator, and $\underline{\mathbf{H}}_{\mathbf{M}}$ is its Jacobian:

$$\begin{aligned}
\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b}_{(i)}) &= \underline{\boldsymbol{\mathcal{H}}}\left(\underline{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b}_{(i)})\right) \\
&= \begin{pmatrix} \boldsymbol{\mathcal{H}}_0(\bullet) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\mathcal{H}}_T(\bullet) \end{pmatrix}\begin{pmatrix} \boldsymbol{\mathcal{M}}_{0,0}(\mathbf{x}^{\rm b}_{(i)}) \\ \vdots \\ \boldsymbol{\mathcal{M}}_{0,T}(\mathbf{x}^{\rm b}_{(i)}) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\mathcal{H}}_0(\mathbf{x}^{\rm b}_{(i)}) \\ \vdots \\ \boldsymbol{\mathcal{H}}_T(\boldsymbol{\mathcal{M}}_{0,T}(\mathbf{x}^{\rm b}_{(i)})) \end{pmatrix},
\end{aligned} \tag{30}$$

$$\begin{aligned}
\underline{\mathbf{H}}_{\mathbf{M}} = \underline{\mathbf{H}}\underline{\mathbf{M}} &= \begin{pmatrix} \mathbf{H}_0 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}_T \end{pmatrix}\begin{pmatrix} \mathbf{M}_{0,0} \\ \vdots \\ \mathbf{M}_{0,T} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_T\mathbf{M}_{0,T} \end{pmatrix},
\end{aligned} \tag{31}$$

where $\boldsymbol{\mathcal{H}}_t(\bullet)$ means use the quantity to the right as an argument and $\boldsymbol{\mathcal{H}}_t$ and $\mathbf{H}_t$ are defined in section 2. The above also serves to define the operators $\underline{\boldsymbol{\mathcal{H}}}$, $\underline{\boldsymbol{\mathcal{M}}}$, $\underline{\mathbf{H}}$ and $\underline{\mathbf{M}}$. Furthermore the linear step $\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}$ can be avoided by approximating with the nonlinear operations $\left\{\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b}_{(i)}) - \underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\overline{\mathbf{x}^{\rm b}})\right\}/\sqrt{N-1}$ (approximating the ith column of $\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}$). This EnKS is ⑤ in Figure 1.

3.3.2. An EnKS with a 4D state vector

An EnKS may be alternatively formulated with 4D state vectors (van Leeuwen and Evensen, 1996), contained in the $(T+1)n \times N$-element matrix $\underline{\mathbf{X}}^{\rm b}$ (the ith column being $\delta\underline{\mathbf{x}}^{\rm b}_{(i)}/\sqrt{N-1}$, where $\delta\underline{\mathbf{x}}^{\rm b}_{(i)}$ is the 4D analogue of (27)):

$$\underline{\mathbf{X}}^{\rm b} = \begin{pmatrix} \mathbf{X}^{\rm b}(0) \\ \vdots \\ \mathbf{X}^{\rm b}(T) \end{pmatrix} = \underline{\mathbf{M}}\mathbf{X}^{\rm b}, \tag{32}$$

and a similar ensemble exists for the analysis, $\underline{\mathbf{X}}^{\rm a}$. An EnKS can be derived by underlining all quantities (e.g. $\mathbf{X}^{\rm b} \rightarrow \underline{\mathbf{X}}^{\rm b}$) in (28):

$$\underline{\mathbf{x}}^{\rm a}_{(i)} = \underline{\mathbf{x}}^{\rm b}_{(i)} + \underline{\mathbf{X}}^{\rm b}(\underline{\mathbf{H}}\underline{\mathbf{X}}^{\rm b})^{\rm T}\left\{\underline{\mathbf{R}} + \underline{\mathbf{H}}\underline{\mathbf{X}}^{\rm b}(\underline{\mathbf{H}}\underline{\mathbf{X}}^{\rm b})^{\rm T}\right\}^{-1}\underline{\mathbf{d}}_{(i)}, \tag{33}$$

where now $\underline{\mathbf{d}}_{(i)} = \underline{\mathbf{y}}^{\rm o}_{(i)} - \underline{\boldsymbol{\mathcal{H}}}(\underline{\mathbf{x}}^{\rm b}_{(i)})$, and $\underline{\boldsymbol{\mathcal{H}}}(\underline{\mathbf{x}}^{\rm b}_{(i)})$ and $\underline{\mathbf{H}}$ are defined in (30) and (31). Again the linear step $\underline{\mathbf{H}}\underline{\mathbf{X}}^{\rm b}$ can be avoided with the approximation $\{\underline{\boldsymbol{\mathcal{H}}}(\underline{\mathbf{x}}^{\rm b}_{(i)}) - \underline{\boldsymbol{\mathcal{H}}}(\overline{\underline{\mathbf{x}}^{\rm b}})\}/\sqrt{N-1}$ (approximating the ith column of $\underline{\mathbf{H}}\underline{\mathbf{X}}^{\rm b}$). This EnKS is ⑥ in Figure 1. An EnKS may also be posed in a formulation which considers observations in batches over the window (Evensen and van Leeuwen, 2000).

The EnKS formulations with 3D and 4D state vectors will be important in the next section where they form the basis of methods that we shall call En4DVar and 4DEnVar respectively.

4. Pure EnVar: combining an ensemble with Var

The last decade has seen experimentation with combining the EnKF with Var. Perhaps the simplest means of doing this is to use a Var analysis to recentre the EnKF analysis ensemble. This (one-way) coupling has been used by Zhang *et al.* (2009) with the simple Lorenz 96 model (Lorenz, 1996), and by Bowler *et al.* (2008) with MOGREPS (the Met Office Global and Regional Ensemble Prediction System).

However, in this section we look at one-way coupling the other way round and review two (related) methods of using the EnKF ensemble forecasts to define the background-error statistics in Var (collectively known as 'EnVar'). We focus on 4D methods (3D methods follow in the same way as in section 2.1.3). The two methods differ in the use – or not – of the linear/adjoint models as used in 4D-Var. Following the nomenclature recommended by Lorenc (2013), we call a pure EnVar system that uses a linear model 'En4DVar' (the labels 'E4DVar', '4DVarBen', and '4DVarBenkf' have also been used by authors). In all such labels, '4DVar' appears, indicating that it uses the same linearized machinery as 4D-Var does. Also following Lorenc (2013), we call a pure EnVar system that does not use a linear model '4DEnVar'. En4DVar and 4DEnVar are discussed below (without localization at this stage). Bear in mind that this nomenclature is not consistent with all the literature, especially in articles prior to Lorenc (2013). Liu *et al.* (2008, 2009) for instance called their method En4DVar even though it does *not* use a linear model, but this has been changed to the recommended 4DEnVar in later articles (e.g. Liu and Xiao, 2013).

Note that *pure* En4DVar and *pure* 4DEnVar use background errors from one source (here the ensemble), but further coupling can be achieved by merging the ensemble statistics with $\mathbf{B}_0$ of Var. This forms another family of methods called '*hybrid* EnVar' (Lorenc, 2013) (section 5) but again this term is not always used consistently in the literature.

4.1. Pure En4DVar: requiring linear/adjoint models

A deterministic analysis $\mathbf{x}^{\rm a}$ (essentially the ensemble mean of (29)) may be found using a Var technique. Consider the cost function (Liu *et al.*, 2008):

$$\begin{aligned}
&J^{\rm En4DVar}(\boldsymbol{\chi}_{\rm ens}) \\
&\quad = \frac{1}{2}\left\|\boldsymbol{\chi}_{\rm ens}\right\|^2_{\mathbf{I}} + \frac{1}{2}\sum_{t=0}^{T}\left\|\mathbf{d}_t - \mathbf{H}_t\mathbf{M}_{0,t}\mathbf{X}^{\rm b}\boldsymbol{\chi}_{\rm ens}\right\|^2_{\mathbf{R}_t^{-1}} &(34)\\
&\quad = \frac{1}{2}\left\|\boldsymbol{\chi}_{\rm ens}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\rm b}\boldsymbol{\chi}_{\rm ens}\right\|^2_{\underline{\mathbf{R}}^{-1}}, &(35)
\end{aligned}$$

which is a function of a new N-element control variable $\boldsymbol{\chi}_{\rm ens}$, and where $\mathbf{d}_t = \mathbf{y}^{\rm o}_t - \boldsymbol{\mathcal{H}}_t\left\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\rm b})\right\}$ and $\underline{\mathbf{d}} = \underline{\mathbf{y}}^{\rm o} - \underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\rm b})$. The latter of the two equations uses the time-compact (underlined)

notation. This control variable comprises coefficients that multiply the ensemble perturbations in the linear combination as follows:

$$\delta\mathbf{x} = \mathbf{X}^{\mathrm{b}}\boldsymbol{\chi}_{\mathrm{ens}} \tag{36}$$

(Lorenc, 2003). Here $\mathbf{X}^{\mathrm{b}}$ acts as a CVT, which is taken from a parallel EnKF system (or even from a population of NMC-style forecast differences (Wang *et al.*, 2014), where the National Meteorological Centre (NMC) method approximates forecast errors with a set of lagged forecasts (Parrish and Derber, 1992)).

To see how (34) works, compare the above equations with those of preconditioned Var (section 2.3). Comparing (34) with (21) (with definition (20) [where in (21) and (20) no model error is considered, $\delta\boldsymbol{\gamma}_{\mathrm{var}}(t) = 0$]). These systems have the same form, except that the CVT $\mathbf{U}$ in (17) becomes $\mathbf{X}^{\mathrm{b}}$ in (36). The implied $\mathbf{B}$-matrix of En4DVar is then akin to (24), i.e. $\mathbf{X}^{\mathrm{b}}\mathbf{X}^{\mathrm{b}\mathrm{T}}$, as consistent with (26). Minimizing (34) is efficient as it is well conditioned, and $\boldsymbol{\chi}_{\mathrm{ens}}$ has only N elements. At the minimum of (34) $\boldsymbol{\chi}_{\mathrm{ens}} = \boldsymbol{\chi}^{\mathrm{a}}_{\mathrm{ens}}$, which leads to the analysis $\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\boldsymbol{\chi}^{\mathrm{a}}_{\mathrm{ens}}$.

The gradient of (34) (Appendix A) is required for the minimization and has the forms:

$$\begin{aligned}
&\nabla_{\boldsymbol{\chi}^{\mathrm{ens}}} J^{\mathrm{En4DVar}} \\
&= \boldsymbol{\chi}_{\mathrm{ens}} - \mathbf{X}^{\mathrm{b}\mathrm{T}} \sum_{t=0}^{T} \mathbf{M}_{0,t}^{\mathrm{T}} \mathbf{H}_t^{\mathrm{T}} \mathbf{R}_t^{-1} \left(\mathbf{d}_t - \mathbf{H}_t \mathbf{M}_{0,t} \mathbf{X}^{\mathrm{b}} \boldsymbol{\chi}_{\mathrm{ens}} \right) \\
&= \boldsymbol{\chi}_{\mathrm{ens}} - \mathbf{X}^{\mathrm{b}\mathrm{T}} \underline{\mathbf{H}}_{\mathbf{M}}^{\mathrm{T}} \underline{\mathbf{R}}^{-1} \left(\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}} \mathbf{X}^{\mathrm{b}} \boldsymbol{\chi}_{\mathrm{ens}} \right),
\end{aligned} \tag{37}$$

where, as in 4D-Var, the linear and adjoint operators appear. This method is formally called En4DVar (⑦ in Figure 1) and the 3D counterpart is called En3DVar (⑧). An accepted alternative name for this method is 4D/3DVarBen (Lorenc, 2013), which translates as 4D/3D-Var with background-error covariance matrix implied from an ensemble. Our recommended nomenclature (broadly in line with Lorenc (2013)) is summarized in Figure 1 (but note the comments at the start of section 4). Minimizing J^{En4DVar} when the model and observation operators are linear is equivalent to the mean EnKS solution in section 3.3.1 (Appendix C).

4.2. Pure 4DEnVar: avoiding linear/adjoint models

Just as the linear and adjoint operators can be avoided in the EnKF and EnKS (section 3), a similar approximation can be applied to En4DVar. For p_t observations at time t, the combination $\mathbf{H}_t\mathbf{M}_{0,t}\mathbf{X}^{\mathrm{b}}$ can be translated to the explicit $p_t \times N$-element matrix $\mathbf{Y}^{\mathbf{x}}_t$ using the nonlinear operators $\boldsymbol{\mathcal{H}}_t$ and $\boldsymbol{\mathcal{M}}_{0,t}$ (Liu *et al.*, 2008):

$$\begin{aligned}
\mathbf{Y}^{\mathbf{x}}_t &= \mathbf{H}_t\mathbf{M}_{0,t}\mathbf{X}^{\mathrm{b}} \\
&\approx \frac{1}{\sqrt{N-1}} \Big(\boldsymbol{\mathcal{H}}_t\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(1)})\} - \overline{\mathbf{y}^{\mathbf{x}}_t}, \cdots \\
&\qquad \cdots, \boldsymbol{\mathcal{H}}_t\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(N)})\} - \overline{\mathbf{y}^{\mathbf{x}}_t} \Big),
\end{aligned} \tag{38}$$

$$\text{and let } \underline{\mathbf{Y}}^{\mathbf{x}} = \begin{pmatrix} \mathbf{Y}^{\mathbf{x}}_0 \\ \vdots \\ \mathbf{Y}^{\mathbf{x}}_T \end{pmatrix} = \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\mathrm{b}}, \tag{39}$$

where $\overline{\mathbf{y}^{\mathbf{x}}_t}$ is the model observation vector at time t based on the ensemble mean, $\overline{\mathbf{y}^{\mathbf{x}}_t} = \boldsymbol{\mathcal{H}}_t\{\boldsymbol{\mathcal{M}}_{0,t}(\overline{\mathbf{x}^{\mathrm{b}}})\}$ and $\underline{\mathbf{Y}}^{\mathbf{x}}$ has dimension $p \times N$. The (time-compact) cost function and gradient of this formulation are adapted from (35) and (37) using (39):

$$J^{\mathrm{4DEnVar}}(\boldsymbol{\chi}_{\mathrm{ens}}) = \frac{1}{2}\left\|\boldsymbol{\chi}_{\mathrm{ens}}\right\|^2_{\mathbf{I}} + \left\|\underline{\mathbf{d}} - \underline{\mathbf{Y}}^{\mathbf{x}}\boldsymbol{\chi}_{\mathrm{ens}}\right\|^2_{\underline{\mathbf{R}}^{-1}}, \tag{40}$$

$$\nabla_{\boldsymbol{\chi}_{\mathrm{ens}}} J^{\mathrm{4DEnVar}} = \boldsymbol{\chi}_{\mathrm{ens}} - \underline{\mathbf{Y}}^{\mathbf{x}\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{Y}}^{\mathbf{x}}\boldsymbol{\chi}_{\mathrm{ens}}\right). \tag{41}$$

Notice that $\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(i)}) = \mathbf{x}^{\mathrm{b}}_{(i)}(t)$ (comprising $\underline{\mathbf{X}}^{\mathrm{b}}$ in (32)) may be computed in advance, which means that this method is akin to using an EnKS with a 4D state vector (section 3.3.2). In (40) and (41), the CVT is $\delta\underline{\mathbf{x}} = \underline{\mathbf{X}}^{\mathrm{b}}\boldsymbol{\chi}_{\mathrm{ens}}$. This method is formally called 4DEnVar (⑨ in Figure 1) (Desroziers *et al.*, 2014; Fairbairn *et al.*, 2014), and is similar to methods of Hunt *et al.* (2004) and Tian *et al.* (2008). The 3D counterpart is 3DEnVar (although it is essentially the same as En3DVar) and alternative names are given in sections 4 and 7.5.

4.3. Comments on EnVar methods

For En4DVar and 4DEnVar the implied background-error covariance at $t = 0$ is $\mathbf{X}^{\mathrm{b}}\mathbf{X}^{\mathrm{b}\mathrm{T}}$, and between times t_1 and t_2 it is $\mathbf{X}^{\mathrm{b}}(t_1)\mathbf{X}^{\mathrm{b}\mathrm{T}}(t_2)$, where for En4DVar $\mathbf{X}^{\mathrm{b}}(t)$ comprises columns of $\mathbf{X}^{\mathrm{b}}$ propagated to time t by $\mathbf{M}_{0,t}$ ($\mathbf{X}^{\mathrm{b}}(t) = \mathbf{M}_{0,t}\mathbf{X}^{\mathrm{b}}$) but for 4DEnVar $\mathbf{X}^{\mathrm{b}}(t)$ comprises columns propagated by the nonlinear model as $\left\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(i)}) - \boldsymbol{\mathcal{M}}_{0,t}(\overline{\mathbf{x}^{\mathrm{b}}})\right\}/\sqrt{N-1}$ (for column i). En4DVar and 4DEnVar are each used typically to find a single deterministic analysis per analysis cycle, and the perturbed ensemble members, $\mathbf{X}^{\mathrm{b}}$, are used in these methods only to define the CVT. There are advantages and disadvantages of EnVar over traditional Var:

- They do not require a background-error covariance model as Var does. This is important when modelling background errors in $\mathbf{B}_0$ involving processes that are too complicated, nonlinear or when geophysical balances are not relevant.
- The control vector $\boldsymbol{\chi}^{\mathrm{ens}}$ has N elements while the control vector $\delta\boldsymbol{\chi}$ in 3D-Var or strong-constraint 4D-Var has n elements. As $N \ll n$, this represents an efficient problem that has shrunk in size to reflect the low rank of $\mathbf{P}^{\mathrm{b}}_{(N)}$.
- 4DEnVar does not need linearized models, but 4D-Var and En4DVar do. If NWP centres no longer have to develop, maintain and run linear and adjoint models, this results in a significant cost gain, especially if linearizing highly nonlinear processes with $\mathbf{M}$ can be avoided (e.g. Xu, 1996; Stiller and Ballard, 2009; Stiller, 2009). As a consequence however, other quantities that rely on the adjoint will no longer be available, such as adjoint sensitivity analysis (e.g. Benedetti *et al.*, 2003) or Hessian singular vector calculations (e.g. Lawrence *et al.*, 2009).
- As with the EnKF/S, EnVar has a suitability for parallel processing for computing the matrix $\underline{\mathbf{Y}}^{\mathbf{x}}$. Such systems have limitless parallelizability with the number of ensemble members.
- The low-rank property of the implied background-error covariance matrix in EnVar means that sampling error problems will inevitably arise when N is small. This usually requires some kind of mitigation such as localization (section 6).

4.4. Generating an ensemble within EnVar

The EnVar forms given in sections 4.1 and 4.2 produce only a single analysis (the mode of the posterior),[¶] while the EnKF/S produce an ensemble of possible analyses. However EnVar permits further constraints to the analysis, such as a tangent linear normal mode constraint (Kleist *et al.*, 2009; Wang *et al.*, 2013) or a variationally-based initialization term applied with an extra term—normally called J_C—added to the cost function

[¶]Instead of a single analysis, such minimization problems can, if required, be repeated for each ensemble member i, e.g. (Liu and Xiao, 2013), where the following substitutions are made in (35): $\mathbf{x}^{\mathrm{b}} \rightarrow \mathbf{x}^{\mathrm{b}}_{(i)}$, $\mathbf{y}^{\mathrm{o}}_t \rightarrow \mathbf{y}^{\mathrm{o}}_{t(i)}$ and $\boldsymbol{\chi}_{\mathrm{ens}} \rightarrow \boldsymbol{\chi}_{\mathrm{ens}(i)}$, where $\mathbf{y}^{\mathrm{o}}_{t(i)}$ is the ith perturbed observation vector, and $\delta\mathbf{x}_{(i)} = \mathbf{X}^{\mathrm{b}}\boldsymbol{\chi}_{\mathrm{ens}(i)}$ is analysis perturbation i. The resulting ensemble of N EnVar analyses is equivalent to (29).

(Clayton *et al.*, 2012; Ge *et al.*, 2012), which can help the analysis to be close to a defined balance.

We also mention a couple of related EnVar schemes that are capable of producing an ensemble. The 'Maximum Likelihood Ensemble Filter' (MLEF) (Zupanski, 2005) can also bypass use of the linearized model. MLEF, which preconditions the variational problem on the Hessian, has been demonstrated in a 3D context. This preconditioning replaces (36) with the following CVT:

$$\delta\mathbf{x} = \mathbf{X}^{\mathrm{b}}(\mathbf{I}+\mathbf{S})^{-\mathrm{T}/2}\boldsymbol{\chi}_{\mathrm{MLEF}}, \tag{42}$$

where in 3D $\mathbf{S} = \left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^{\mathrm{b}}\right)^{\mathrm{T}}\left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^{\mathrm{b}}\right)$, which is written in a way so that the $\mathbf{H}$ (as before) can be approximated with two runs of the nonlinear observation operator. Hessian preconditioning not only allows a very efficient minimization but it also allows calculation of an analysis ensemble, so the method can be used without a separate ensemble generator. The 'Ensemble Variational Integrated localized (or Lanczos)' (EVIL) scheme of Auligné *et al.* (2016) is another method (discussed further in section 5.5 where it is applied to a hybrid setting).

5. Hybrid methods

According to the definition in Lorenc (2013), a hybrid method refers to one that blends $\mathbf{B}_0$ from Var with $\mathbf{P}^{\mathrm{b}}_{(N)}$ from an ensemble (recall $\mathbf{B}_0$ is full-rank but quasi-static and crudely modelled, and $\mathbf{P}^{\mathrm{b}}_{(N)}$ is flow-dependent but rank deficient). Six ways of combining these matrices are discussed here. The first is an ensemble-based recalibration of the $\mathbf{B}_0$-matrix in Var; the second and third are (equivalent) ways of averaging $\mathbf{B}_0$ and $\mathbf{P}^{\mathrm{b}}_{(N)}$; the fourth averages the gain matrices instead of the covariance matrices; the fifth returns to averaging $\mathbf{B}_0$ and $\mathbf{P}^{\mathrm{b}}_{(N)}$ but provides an ensemble of analyses as part of the variational iterations; and the sixth combines $\mathbf{B}_0$ with $\mathbf{P}^{\mathrm{b}}_{(N)}$, but not by averaging them.

Hybridization introduces climatological error information into EnsVar (where the hybrid system becomes less sensitive to ensemble size than the pure EnKF; Zhang *et al.*, 2009). The use of hybrid covariances has long been thought to be useful (Gustafsson, 2007; Bishop and Satterfield, 2013). Here we shall consider the former viewpoint (equivalent to hybridization of, for example, 4D-Var with En4DVar). From this perspective, the hybrid background-error covariance matrix, here denoted $\mathbf{B}^{\mathrm{h}}$, replaces the usual $\mathbf{B}_0$-matrix in Var. Localization is not discussed until section 6 since at this stage it is a distraction from the hybridization procedure.

5.1. *Ensemble recalibration of* $\mathbf{B}_0$ *in Var*

Some operational centres use a parallel ensemble prediction system to continuously recalibrate the model of $\mathbf{B}_0$ in their Var system (this blending of ensemble and climatological information would arguably classify this as a hybrid method, so we should call it a $\mathbf{B}^{\mathrm{h}}$-matrix). Météo-France and the ECMWF, for example, use the spread of their ensemble prediction systems (initialized not from an EnKF but from ensembles of perturbed 4D-Var assimilations; Belo Pereira and Berre, 2006; Berre *et al.*, 2007, 2009; Isaksen *et al.*, 2010) to model the 'variances of the day' of $\mathbf{B}^{\mathrm{h}}$ (Bonavita *et al.*, 2011, 2012; Raynaud *et al.*, 2012). The ensemble is known to have significant sampling errors (and so needs to be filtered) and is known to be under-spread (and so needs to be inflated).

Recently the ECMWF and Météo-France extended their hybrid schemes to include on-line calibration of the spatial correlations in their $\mathbf{B}$-matrices. At the ECMWF, e.g. these spatial structures are described by a 'wavelet diagonal' scheme (Fisher, 2003), which can represent different correlation length-scales for different geographical regions. The original implementation of this scheme used a climatological calibration, which implied static background-error length-scales, but these are known in reality to be flow-dependent (e.g. longer length-scales in higher pressure systems). The ECMWF has experimented with two ways of calibrating their wavelet scheme, namely using samples of background errors from either (i) a 12 day window (25 ensemble members, twice per day, leading to 600 samples), or (ii) a combination of current members (25 members within an 8 h window – 200 samples) plus samples representing climatology (400 samples spread over the seasons, leading again to 600 samples in total) (Bonavita *et al.*, 2016). The 600 samples in each case are considered necessary to calibrate the wavelet model without filtering or localizing. Even though the forms of their $\mathbf{B}^{\mathrm{h}}$-matrix models (Derber and Bouttier, 1999; Fisher, 2003) are not changed by recalibration, important flow-dependence is added as a consequence of recalibration with either (i) or (ii) above, resulting in modestly improved forecasts right out to 10 days in some quantities (Bonavita *et al.*, 2016).

5.2. *The explicit average of error covariance matrices*

The classic hybrid background-error covariance matrix, $\mathbf{B}^{\mathrm{h}}$, is defined according to Hammill and Snyder (2000) as the weighted average of $\mathbf{B}_0$ and $\mathbf{P}^{\mathrm{b}}_{(N)}$:

$$\mathbf{B}^{\mathrm{h}} = (1-\beta)\mathbf{B}_0 + \beta\mathbf{P}^{\mathrm{b}}_{(N)}, \tag{43}$$

where β ($0 \le \beta \le 1$) is a tunable factor controlling the weighting of the static and flow-dependent covariances, and $\mathbf{B}^{\mathrm{h}}$ replaces $\mathbf{B}_0$ in (10). The use of explicit matrices is obviously impractical for large systems and so methods of using $\mathbf{B}^{\mathrm{h}}$ implicitly are introduced.

5.3. *The implicit average of error covariance matrices*

This method merges the CVTs of (17) and (36) to represent $\mathbf{B}_0$ and $\mathbf{P}^{\mathrm{b}}_{(N)}$ implicitly. We show this in the context of hybridizing 4D-Var with En4DVar:

$$J^{\mathrm{HEn4DVar}}(\delta\boldsymbol{\chi}_{\mathrm{var}}, \boldsymbol{\chi}_{\mathrm{ens}}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\boldsymbol{\chi}_{\mathrm{ens}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\sum_{t=0}^{T}\|\delta\mathbf{d}_t\|^2_{\mathbf{R}_t^{-1}}, \tag{44}$$

where

$$\delta\mathbf{d}_t = \mathbf{d}_t - \mathbf{H}_t\mathbf{M}_{0,t}\left(\sqrt{1-\beta}\,\mathbf{U}\delta\boldsymbol{\chi}_{\mathrm{var}} + \sqrt{\beta}\,\mathbf{X}^{\mathrm{b}}\boldsymbol{\chi}_{\mathrm{ens}}\right) \tag{45}$$

$$\text{and } \mathbf{d}_t = \mathbf{y}^{\mathrm{o}}_t - \boldsymbol{H}_t\left\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}})\right\},$$

(cf. (44)–(45) with (34)). In (44) the augmented control vector $(\delta\boldsymbol{\chi}_{\mathrm{var}}, \boldsymbol{\chi}_{\mathrm{ens}})$ has $n+N$ elements where $\delta\boldsymbol{\chi}_{\mathrm{var}}$ is associated with $\mathbf{B}_0$ (as $\delta\boldsymbol{\chi}_{\mathrm{var}}$ in section 2.3), and $\boldsymbol{\chi}_{\mathrm{ens}}$ associated with $\mathbf{P}^{\mathrm{b}}_{(N)}$ (as $\boldsymbol{\chi}_{\mathrm{ens}}$ in section 4). Recall that n is the size of the state, N is the number of ensemble members, $\mathbf{U}$ is the original $n\times n$ CVT of section 2.3, and $\mathbf{X}^{\mathrm{b}}$ is the $n\times N$ matrix comprising the ensemble perturbations $(\mathbf{x}^{\mathrm{b}}_{(i)} - \overline{\mathbf{x}^{\mathrm{b}}})/\sqrt{N-1}$. The hybrid En4DVar cost function can be compactly written as

$$J^{\mathrm{HEn4DVar}}(\boldsymbol{\chi}_{\mathrm{h}}) = \frac{1}{2}\left\|\boldsymbol{\chi}_{\mathrm{h}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\sum_{t=0}^{T}\|\delta\mathbf{d}_t\|^2_{\mathbf{R}_t^{-1}}, \tag{46}$$

$$\text{where } \boldsymbol{\chi}_{\mathrm{h}} = \begin{pmatrix}\delta\boldsymbol{\chi}_{\mathrm{var}}\\ \boldsymbol{\chi}_{\mathrm{ens}}\end{pmatrix}. \tag{47}$$

For conciseness, the formula for the gradient of J^{HEn4DVar} with respect to $\boldsymbol{\chi}_{\mathrm{h}}$ is left to section 7, where localization is also

considered. The CVT for (46) relates $\delta\boldsymbol{\chi}_{\rm var}$ and $\boldsymbol{\chi}_{\rm ens}$ to $\delta\mathbf{x}$ at $t=0$:

$$\begin{aligned}\delta\mathbf{x} &= \sqrt{1-\beta}\mathbf{U}\delta\boldsymbol{\chi}_{\rm var} + \sqrt{\beta}\mathbf{X}^{\rm b}\boldsymbol{\chi}_{\rm ens},\\ &= \mathbf{U}_{\rm h}\boldsymbol{\chi}_{\rm h} = \left(\sqrt{1-\beta}\mathbf{U} \quad \sqrt{\beta}\mathbf{X}^{\rm b}\right)\boldsymbol{\chi}^{\rm h},\end{aligned} \tag{48}$$

(cf. (17)). The hybrid CVT which consolidates $\mathbf{U}$ and $\mathbf{X}^{\rm b}$ is called $\mathbf{U}_{\rm h}$, which is the $n\times(n+N)$ matrix $\left(\sqrt{1-\beta}\mathbf{U} \quad \sqrt{\beta}\mathbf{X}^{\rm b}\right)$. In the $\boldsymbol{\chi}_{\rm h}$ representation, background errors have covariance $\left\langle\boldsymbol{\chi}^{\rm h}\boldsymbol{\chi}^{\rm h\,T}\right\rangle_{\rm b} = \mathbf{I}$, which implies the following error covariance in the model representation ($\delta\mathbf{x}$):

$$\begin{aligned}\mathbf{B}_{\rm h} &= \left\langle\delta\mathbf{x}\delta\mathbf{x}^{\rm T}\right\rangle_{\rm b}\\ &= \left(\sqrt{1-\beta}\mathbf{U} \quad \sqrt{\beta}\mathbf{X}^{\rm b}\right)\left\langle\boldsymbol{\chi}^{\rm h}\boldsymbol{\chi}^{\rm h\,T}\right\rangle_{\rm b}\begin{pmatrix}\sqrt{1-\beta}\mathbf{U}^{\rm T}\\ \sqrt{\beta}\mathbf{X}^{\rm b\,T}\end{pmatrix}\\ &= (1-\beta)\mathbf{U}\mathbf{U}^{\rm T}+\beta\mathbf{X}^{\rm b}\mathbf{X}^{\rm b\,T} = (1-\beta)\mathbf{B}_0+\beta\mathbf{P}^{\rm b}_{(N)},\end{aligned} \tag{49}$$

where $\langle\bullet\rangle_{\rm b}$ performs an expectation over the background population. Equations (49) and (43) are identical, which shows that minimizing (44) with respect to $\delta\boldsymbol{\chi}_{\rm h}$ (with CVT (48)), and then transforming the resulting control variable increment into model space with (48), is the same as minimizing the strong-constraint 4D-Var cost function with $\mathbf{B}_0 \rightarrow \mathbf{B}_{\rm h}$.[‖] This is hybrid En4DVar (⑪ in Figure 1, ⑫ for En3DVar).

A benefit of form (44) over (43) is efficiency as it avoids prohibitively large matrices. Wang *et al.* (2007b) provide an alternative derivation of (49). Although this hybrid method of representing the background-error covariance matrix has been applied to strong-constraint 4D-Var, it may also be used with the weak-constraint formulation (21). When $\beta=1$ ($\beta=0$), this system is identical to pure En4DVar (pure Var) (section (4.1)).

5.4. *Combining gain matrices*

The most common type of hybrid schemes average $\mathbf{B}_0$ with $\mathbf{P}^{\rm b}_{(N)}$, but an alternative is to follow Penny (2014) by averaging the Kalman gains of pure Var (gain $\mathbf{K}^{\rm var}$) and of pure EnKF (gain $\mathbf{K}^{\rm ens}$) to give $\mathbf{K}^{\rm h}$. For 3D schemes, e.g.

$$\mathbf{K}^{\rm h} = \beta_1\mathbf{K}^{\rm ens}+\beta_2\mathbf{K}^{\rm var}+\beta_3\mathbf{K}^{\rm var}\mathbf{H}\mathbf{K}^{\rm ens}, \tag{50}$$

$$\text{where } \mathbf{K}^{\rm ens} = \mathbf{P}^{\rm b}_{(N)}\mathbf{H}^{\rm T}\left(\mathbf{H}\mathbf{P}^{\rm b}_{(N)}\mathbf{H}^{\rm T}+\mathbf{R}\right)^{-1}, \tag{51}$$

$$\mathbf{K}^{\rm var} = \mathbf{B}_0\mathbf{H}^{\rm T}\left(\mathbf{H}\mathbf{B}_0\mathbf{H}^{\rm T}+\mathbf{R}\right)^{-1}, \tag{52}$$

and β_i are scalars. $\mathbf{K}^{\rm ens}$ and $\mathbf{K}^{\rm var}$ capture the way that the respective schemes work but are not computed explicitly.

Penny (2014) applied this averaging to 3D-Var and an LETKF with the Lorenz 96 model (Lorenz, 1996) and the choices $\beta_1=1$, $\beta_2=\alpha$, and $\beta_3=-\alpha$. This leads to the hybrid gain:

$$\mathbf{K}^{\rm h} = \mathbf{K}^{\rm ens}+\alpha\mathbf{K}^{\rm var}(\mathbf{I}-\mathbf{H}\mathbf{K}^{\rm ens}), \tag{53}$$

and the hybrid analysis $\mathbf{x}^{\rm a} = \mathbf{x}^{\rm b}+\mathbf{K}^{\rm h}(\mathbf{y}^{\rm o}-\mathbf{H}\mathbf{x}^{\rm b})$. This analysis is equivalent to first running the EnKF to give the ensemble mean $\bar{\mathbf{x}}^{\rm a}_{\rm ens}$, running Var with this as the background to give $\mathbf{x}^{\rm a}_{\rm var}$, and then doing the following weighted average to give the hybrid analysis $\mathbf{x}^{\rm a}_{\rm h} = \bar{\mathbf{x}}^{\rm a}_{\rm ens}+\alpha\mathbf{x}^{\rm a}_{\rm var}$. Penny (2014) found that this kind of hybrid improves the analysis over those of the separate pure schemes. However the benefit of this hybrid is that it requires little extra coding to existing EnKF and Var schemes.

[‖]It is possible to allow the β to vary with vertical level, as is done by Buehner *et al.* (2013). Some authors put the $\sqrt{1-\beta}$ and $\sqrt{\beta}$ terms on the ensemble and climatological terms respectively instead of the way round in (48). Other authors (e.g. Wang *et al.*, 2008b; Gustafsson *et al.*, 2014), prefer to set the β weights in the cost function instead of in the CVT. In this case the background 'Var' and 'ens' terms would contain the extra factors $1/(1-\beta)$ and $1/\beta$ respectively and the $\sqrt{1-\beta}$ and $\sqrt{\beta}$ would be omitted in (48).

5.5. *EVIL*

Most EnVar systems use a separate EnKF-style system to generate $\mathbf{X}^{\rm b}$ for each cycle. The fact that the EnVar and EnKF are separate can lead to inconsistencies, and hence to sub-optimality in hybrid EnVar. EVIL (Ensemble Variational Integrated Localized [or Lanzos]) (Auligné *et al.*, 2016) represents a modification to hybrid EnVar which uses information gained in the Var minimization procedure to estimate an analysis ensemble, instead of the usual single analysis.

EVIL is based on the fact that conjugate gradient (CG)-based minimization algorithms are closely related to Lanczos methods (Paige and Saunders, 1975; Fisher and Courtier, 1995; El Akkraoui *et al.*, 2013). In outline, the gradient descent vectors from q iterations of the CG procedure form a Krylov sub-space in which the Hessian of the cost function is tridiagonal (q here takes the role of N). Typically $q\sim\mathcal{O}(50)$, so such a Hessian can be easily diagonalized. The eigenvectors of the Hessian in control space – stored in the $n\times q$ matrix $\mathbf{Z}_q$ - and the eigenvectors - in the $q\times q$ diagonal matrix $\boldsymbol{\Theta}$ – are called Ritz pairs. In general in quadratic systems, the analysis-error covariance matrix, $\mathbf{A}$, is the inverse of the Hessian. This leads to the following approximation for an $n\times q$ analysis ensemble, $\mathbf{X}^{\rm a}$ in 3D (Auligné *et al.*, 2016):

$$\begin{aligned}\mathbf{X}^{\rm a} &= \mathbf{X}^{\rm b}+\mathbf{A}\mathbf{H}^{\rm T}\mathbf{R}^{-1}\left(\mathbf{Y}^{\rm o}-\mathbf{H}\mathbf{X}^{\rm b}\right),\\ &\approx \mathbf{X}^{\rm b}+\mathbf{U}_{\rm h}\mathbf{Z}_q\boldsymbol{\Theta}_q^{-1}\mathbf{Z}_q^{\rm T}\mathbf{U}_{\rm h}^{\rm T}\mathbf{H}^{\rm T}\mathbf{R}^{-1}\left(\mathbf{Y}^{\rm o}-\mathbf{H}\mathbf{X}^{\rm b}\right),\end{aligned} \tag{54}$$

where $\mathbf{Z}_q\boldsymbol{\Theta}_q^{-1}\mathbf{Z}_q^{\rm T}$ is the analysis-error covariance (in control space) found from the CG/Lanczos procedure, $\mathbf{Y}^{\rm o}$ is a $p\times q$ matrix of stochastically perturbed observations, and $\mathbf{U}_{\rm h}$ is the hybrid CVT (48). (Note that the 3D objects like $\mathbf{X}^{\rm b}$, $\mathbf{H}$, etc. could be extended to 4D with the underline notation of sections 3.3.1 and 3.3.2 as appropriate.) The EnVar minimization would provide the Ritz pair information and (54) would provide the machinery to compute an analysis ensemble – then propagated in time for the next forecast ensemble. EVIL is ⑩ in Figure 1.

Equation (54) is a particular implementation of EVIL called stochastic-EVIL (S-EVIL). Other implementations are deterministic EVIL (D-EVIL), which uses the Ritz pairs to compute a square-root of the analysis-error covariance matrix, and a resampling EVIL (R-EVIL) which can yield additional ensemble members (Auligné *et al.*, 2016). It may also be applied to bi-conjugate gradient algorithms giving dual representations of the Ritz pairs.

The number of iterations, q, may need to be relatively high to obtain a reasonable representation of the correct eigenspectrum of $\mathbf{A}$, and Auligné *et al.* (2016) suggest that several hundred may be needed in their cut-down (single-level) system. This may not yet be affordable to run routinely, but is an option for future systems. The ability of $\mathbf{Z}_q\boldsymbol{\Theta}_q^{-1}\mathbf{Z}_q^{\rm T}$ to represent the analysis-error covariance matrix will need to be checked thoroughly, as will the performance of the method with nonlinear operators where the link between the Hessian and the analysis-error covariance matrix is no longer strictly valid.

5.6. *The ensemble reduced-rank Kalman filter*

In (43), $\mathbf{B}_0$ is able to influence how the DA affects all degrees of freedom, including the subspace spanned by the ensemble. In Petrie and Bannister (2011), another hybrid scheme is proposed which, instead of averaging, essentially uses $\mathbf{P}^{\rm b}_{(N)}$ in the ensemble subspace and uses $\mathbf{B}_0$ elsewhere. The general method is based on the 'reduced-rank Kalman filter' (or the 'simplified Kalman filter') developed at ECMWF in the 1990s for use with Hessian singular vectors (Fisher, 1998; Fisher and Andersson, 2001; Beck and Ehrendorfer, 2005), but Petrie and Bannister (2011) adapted it for use with an ensemble and called it the 'Ensemble Reduced-Rank Kalman Filter' (EnRRKF).

The cost function for the EnRRKF is a function of the n-element control vector $\boldsymbol{\chi}_{\text{EnRRKF}}$:

$$\begin{aligned} J^{\text{EnRRKF}}(\boldsymbol{\chi}_{\text{EnRRKF}}) &= \left\| \boldsymbol{\chi}_{\text{EnRRKF}} \right\|^2_{\mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}}} \\ &= \frac{1}{2} \left\| \boldsymbol{\chi}_{\text{s}} \right\|^2_{\mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}}} + \bar{\boldsymbol{\chi}}_{\text{s}}^{\text{T}} \mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}} \boldsymbol{\chi}_{\text{s}} \\ &\quad + \frac{1}{2} \left\| \bar{\boldsymbol{\chi}}_{\text{s}} \right\|^2_{\mathbf{I}} + \frac{1}{2} \sum_{t=0}^{T} \left\| \delta \mathbf{d}_t \right\|^2_{\mathbf{R}_t^{-1}}, \end{aligned} \tag{55}$$

$$\text{where } \boldsymbol{\chi}_{\text{EnRRKF}} = \boldsymbol{\chi}_{\text{s}} + \bar{\boldsymbol{\chi}}_{\text{s}}, \tag{56}$$

$$\delta \mathbf{d}_t = \mathbf{d}_t - \mathbf{H}_t \mathbf{M}_{0,t} \mathbf{U} \mathbf{X} \boldsymbol{\chi}_{\text{EnRRKF}},$$

$$\text{and } \mathbf{d}_t = \mathbf{y}^{\text{o}}_t - \boldsymbol{\mathcal{H}}_t \left\{ \mathbf{x}^{\text{b}}(t) \right\}.$$

The first part of $\boldsymbol{\chi}_{\text{EnRRKF}}$ in (56) is $\boldsymbol{\chi}_{\text{s}}$ (n-elements) which represents the subspace (the part that has flow-dependent covariances) and is non-zero only in the first N elements. The second part is $\bar{\boldsymbol{\chi}}_{\text{s}}$ (n-elements) which is the remainder (the part that has climatological covariances) and is non-zero only in the last $n-N$ elements. Due to the zero terms, summing these vectors in (56) does not lose any information. The first line of (55) has three terms due to errors that lie
(i) exclusively inside the N-dimensional subspace,
(ii) inside and outside this subspace, and
(iii) exclusively outside this subspace respectively.
The CVT for the EnRRKF is:

$$\delta \mathbf{x} = \mathbf{U} \mathbf{X} \boldsymbol{\chi}_{\text{EnRRKF}}, \tag{57}$$

where $\mathbf{U}$ is the same CVT used in pure Var (17) and $\mathbf{X}$ is a special $n \times n$ matrix ($\mathbf{X}\mathbf{X}^{\text{T}} = \mathbf{I}$ and $\mathbf{X}^{\text{T}}\mathbf{X} = \mathbf{I}$) (do not confuse $\mathbf{X}$ with $\mathbf{X}^{\text{b}}$). $\mathbf{X}$ has the following special properties: when $\delta\mathbf{x}$ lies in the ensemble subspace, $\mathbf{X}^{-1}\mathbf{U}^{-1}\delta\mathbf{x}$ is a vector which can be non-zero only in the first N elements (the last $n-N$ elements are zero), and when $\delta\mathbf{x}$ lies outside the ensemble subspace, $\mathbf{X}^{-1}\mathbf{U}^{-1}\delta\mathbf{x}$ is a vector which can be non-zero only in the last $n-N$ elements (the first N elements are zero). This reflects the structures of $\boldsymbol{\chi}_{\text{s}}$ and $\bar{\boldsymbol{\chi}}_{\text{s}}$ and can be achieved using Householder transforms (Fisher, 1998). Although $\mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}}$ appears in (55), it acts only on vectors that are always zero in the last $n-N$ elements. This means that only part of the inverse matrix $\mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}}$ (the left-most $n \times N$ part) needs to be known, which is achievable for $N \sim \mathcal{O}(10)$–$\mathcal{O}(10^2)$. This reduced $\mathbf{P}^{\text{b}\,-1}_{\boldsymbol{\chi}}$ matrix is derived in Petrie and Bannister (2011) and has a complicated, but calculable form.

To our knowledge, the EnRRKF has not yet been tested, so it is impossible to assess it. Since the flow-dependent part of the error covariance is nearly full-rank in the N-dimensional space (or at least a K-dimensional subspace ($K < N$) can be defined where $\mathbf{P}^{\text{b}}_{(K)}$ is full rank), there is reason to suppose that localization in the EnRRKF is not essential. Unlike for the other hybrid methods, the length of the EnRRKF control vector is no larger than that used for pure Var. Note that the implied background-error covariance matrix for the EnRRKF is not of the form $\mathbf{U}\mathbf{X}(\mathbf{U}\mathbf{X})^{\text{T}}$ as the $\boldsymbol{\chi}_{\text{EnRRKF}}$ does not have uncorrelated background-error covariances.

6. Sampling noise and localization

In ensemble applications in NWP, $N \ll n$, where n is $\geq \mathcal{O}(10^7)$, but N is typically $< \mathcal{O}(10^3)$. This means that $\mathbf{P}^{\text{b}}_{(N)}$ (26) is under-sampled. The consequences of under-sampling are two-fold: $\mathbf{P}^{\text{b}}_{(N)}$ usually underestimates the variance (leading to filter divergence) and $\mathbf{P}^{\text{b}}_{(N)}$ is severely rank deficient; the rank of $\mathbf{P}^{\text{b}}_{(N)}$ computed with (26) is $\leq (N-1)$, which leads to the introduction of analysis noise (Houtekamer and Mitchell, 1998; van Leeuwen, 1999; Hamill *et al.*, 2001; Houtekamer and Mitchell, 2005; Ehrendorfer, 2007; Meng and Zhang, 2012; Houtekamer and Zhang, 2016). In normal situations when $N \ll n$, filter divergence can be mitigated with inflation (Anderson and Anderson, 1999) and rank deficiency can be mitigated with localization (Hamill *et al.*, 2001; Lorenc, 2003).

Localization plays a fundamental role in EnKF-based systems, but before this section, for simplicity, it has been missing from the equations. This section is about how localization is included.

Different means of reducing far-field sampling errors by localization are used. One common method modifies the sample covariances by a Schur product in model space (Lorenc, 2003; Hamill *et al.*, 2001). This is the type of localization that is reviewed below. Another method works in observation space by limiting the observations used to update a particular point to those within a prescribed radius, and/or by gradually reducing the influence of the observations as a function of distance from the point (Houtekamer and Mitchell, 2005). The latter method is used for instance in the local ensemble Kalman filter (Ott *et al.*, 2004; Szunyogh *et al.*, 2005) and with other ensemble DA techniques (Bishop *et al.*, 2001; Pham, 2001; Szunyogh *et al.*, 2008), but presents problems with interpreting observations with non-local observation operators (Fertig *et al.*, 2008), such as satellite radiance measurements.

6.1. Schur product localization applied in model space

Sampling noise due to small N is most significant when the true correlation values are small. Since true correlation values are expected to be small between points that are separated by a large distance, a common approach is to filter computed correlation values according to the distance separating the points. Consider a single field and let the sample covariance between points p and q (at positions $\mathbf{r}_p$ and $\mathbf{r}_q$ respectively) be the matrix element $\left[\mathbf{P}^{\text{b}}_{(N)}\right]_{pq}$. This is multiplied by a moderation function $c(\mathbf{r}_p, \mathbf{r}_q)$ which is unity when $\mathbf{r}_p = \mathbf{r}_q$ and goes to zero when $|\mathbf{r}_p - \mathbf{r}_q| \rightarrow \infty$ (Gaspari and Cohn, 1999). This is localization. In the context of (26), the localized covariance function is

$$c(\mathbf{r}_p, \mathbf{r}_q) \left[\mathbf{P}^{\text{b}}_{(N)}\right]_{pq} = \frac{c(\mathbf{r}_p, \mathbf{r}_q)}{N-1} \sum_{i=1}^{N} \delta \mathbf{x}^{\text{b}}_{(i)}(\mathbf{r}_p) \delta \mathbf{x}^{\text{b}}_{(i)}(\mathbf{r}_q), \tag{58}$$

where $\delta \mathbf{x}^{\text{b}}_{(i)}(\mathbf{r}_p)$ is the ith background perturbation (27) at $\mathbf{r}_p$. Define an $n \times n$ localization matrix $\mathbf{C}$ to have elements $\mathbf{C}_{pq} = c(\mathbf{r}_p, \mathbf{r}_q)$ and define the localized background-error covariance matrix to be $\widehat{\mathbf{P}}^{\text{b}}_{(N)}$, whose matrix elements are $\left[\widehat{\mathbf{P}}^{\text{b}}_{(N)}\right]_{pq} = c(\mathbf{r}_p, \mathbf{r}_q) \left[\mathbf{P}^{\text{b}}_{(N)}\right]_{pq}$. This can be written in the following compact way:

$$\widehat{\mathbf{P}}^{\text{b}}_{(N)} = \mathbf{C} \circ \mathbf{P}^{\text{b}}_{(N)}, \tag{59}$$

where the $\circ$ symbol is the Schur (element-by-element or Hadamard) product, and where a covariance with a hat indicates its localized form. (This hat notation will apply to any quantity: covariance matrices and cost functions, etc.) This localization can be generalized to the multivariate situation (Bishop and Hodyss, 2009a,2009b,2011; Bannister, 2015). Matrix $\mathbf{C}$ is mathematically a correlation matrix, and should be chosen such that $\widehat{\mathbf{P}}^{\text{b}}_{(N)}$ remains a true covariance (Gaspari and Cohn, 1999).

The Schur product localization discussed above may be written implicitly, which allows localization to be used in DA without the need for the explicit covariance matrices in (59). The two known forms are labelled 'B05' and 'L03' after Buehner (2005) and Lorenc (2003) respectively (as is done in Wang *et al.*, 2007b).

6.2. The 'B05' representation of the localization Schur product

In a similar way to the decomposition of $\mathbf{P}^{\text{b}}_{(N)}$ as $\mathbf{P}^{\text{b}}_{(N)} = \mathbf{X}^{\text{b}}\mathbf{X}^{\text{b}\,\text{T}}$ (26), suppose similarly that $\mathbf{C}$ can be decomposed as $\mathbf{C} = \mathbf{U}^{\text{C}}\mathbf{U}^{\text{C}\,\text{T}}$

(where the $n\times M$ matrix $\mathbf{U}^{\mathbf{C}}$ is a kind of 'square-root' of $\mathbf{C}$), and $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}$ can be decomposed as $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}=\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}$ (where the $n\times NM$ matrix $\widehat{\mathbf{X}}^{\mathrm{b}}$ is defined below). Appendix D shows that $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}$ can be written as:

$$\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}=\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}},\ \text{where} \tag{60}$$

$$\begin{aligned}\widehat{\mathbf{X}}^{\mathrm{b}}&=\frac{1}{\sqrt{N-1}}\Big(\delta\mathbf{x}^{\mathrm{b}}_{(1)}\circ\mathbf{u}^{\mathbf{C}}_{(1)},\ldots,\delta\mathbf{x}^{\mathrm{b}}_{(1)}\circ\mathbf{u}^{\mathbf{C}}_{(M)},\ldots\\&\qquad\ldots,\delta\mathbf{x}^{\mathrm{b}}_{(i)}\circ\mathbf{u}^{\mathbf{C}}_{(1)},\ldots,\delta\mathbf{x}^{\mathrm{b}}_{(i)}\circ\mathbf{u}^{\mathbf{C}}_{(M)},\ldots\\&\qquad\ldots,\delta\mathbf{x}^{\mathrm{b}}_{(N)}\circ\mathbf{u}^{\mathbf{C}}_{(1)},\ldots,\delta\mathbf{x}^{\mathrm{b}}_{(N)}\circ\mathbf{u}^{\mathbf{C}}_{(M)}\Big)\\&=\frac{1}{\sqrt{N-1}}\Big(\mathrm{diag}(\delta\mathbf{x}^{\mathrm{b}}_{(1)})\mathbf{U}^{\mathbf{C}},\ldots,\mathrm{diag}(\delta\mathbf{x}^{\mathrm{b}}_{(N)})\mathbf{U}^{\mathbf{C}}\Big)\end{aligned} \tag{61}$$

(Buehner, 2005), where $\delta\mathbf{x}^{\mathrm{b}}_{(i)}/\sqrt{N-1}$ is the vector occupying the ith column of $\mathbf{X}^{\mathrm{b}}$, $\mathbf{u}^{\mathbf{C}}_{(i)}$ is the vector occupying the ith column of $\mathbf{U}^{\mathbf{C}}$, and the diag operator evaluates to an $n\times n$ diagonal matrix whose diagonal comprises its n-element vector argument. In other words, the columns of $\widehat{\mathbf{X}}^{\mathrm{b}}$ are formed in (61) from every possible Schur-product-pair of columns of $\mathbf{X}^{\mathrm{b}}$ and $\mathbf{U}^{\mathbf{C}}$, and $\widehat{\mathbf{X}}^{\mathrm{b}}$ can be thought of as a matrix of NM scaled effective ensemble members whose covariance $\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}$ is the localized background-error covariance $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}$. As $NM>N$, $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}$ would be expected to have a higher rank (maximum rank $\sim NM$) than $\mathbf{P}^{\mathrm{b}}_{(N)}$ (maximum rank $\sim N$). This approach may be used to localize any of the pure EnVar or hybrid schemes discussed in sections 4 and 5 by changing (36) to

$$\delta\mathbf{x}=\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}, \tag{62}$$

where $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$ is the corresponding NM-element control vector.

6.3. *The 'L03' representation of the localization Schur product*

Another compact representation of a localized ensemble background-error covariance matrix is often used for the variational formulations of sections 4 and 5 (Lorenc, 2003). This representation is explained with reference to localizing the En4DVar system (34) (⑦ in Figure 1), which involves N new control vectors:

$$\begin{aligned}\widehat{J}^{\mathrm{En4DVar}}(\widehat{\mathbf{A}})=&\frac{1}{2}\sum_{i=1}^{N}\left\|\widehat{\boldsymbol{\alpha}}_{(i)}\right\|^2_{\mathbf{C}^{-1}}\\&+\frac{1}{2}\sum_{t=0}^{T}\left\|\mathbf{d}_t-\mathbf{H}_t\mathbf{M}_{0,t}\left[\mathbf{X}^{\mathrm{b}}\circ\widehat{\mathbf{A}}\right]\mathbf{1}_N\right\|^2_{\mathbf{R}_t^{-1}},\end{aligned} \tag{63}$$

$$\text{where }\widehat{\mathbf{A}}=\begin{pmatrix}\widehat{\boldsymbol{\alpha}}_{(1)}&\cdots&\widehat{\boldsymbol{\alpha}}_{(N)}\end{pmatrix}, \tag{64}$$

$$\text{and }\delta\mathbf{x}=\frac{1}{\sqrt{N-1}}\sum_{i=1}^{N}\delta\mathbf{x}^{\mathrm{b}}_{(i)}\circ\widehat{\boldsymbol{\alpha}}_{(i)}=\left[\mathbf{X}^{\mathrm{b}}\circ\widehat{\mathbf{A}}\right]\mathbf{1}_N, \tag{65}$$

and $\widehat{\boldsymbol{\alpha}}_{(i)}$ is the n-element control vector associated with the ith ensemble member, which are assembled into the $n\times N$ matrix $\widehat{\mathbf{A}}$, and $\mathbf{1}_N$ is the column vector whose N elements all contain 1. Equation (65) is an unpreconditioned form of L03 localization. A preconditioned form can be devised with N alternative vectors $\widehat{\boldsymbol{\chi}}_{(1)}$ to $\widehat{\boldsymbol{\chi}}_{(N)}$ assembled into the matrix $\widehat{\mathbf{V}}$:

$$\begin{aligned}\widehat{J}^{\mathrm{En4DVar}}(\widehat{\mathbf{V}})=&\frac{1}{2}\sum_{i=1}^{N}\left\|\widehat{\boldsymbol{\chi}}_{(i)}\right\|^2_{\mathbf{I}}\\&+\frac{1}{2}\sum_{t=0}^{T}\left\|\mathbf{d}_t-\mathbf{H}_t\mathbf{M}_{0,t}\left\{\mathbf{X}^{\mathrm{b}}\circ\left(\mathbf{U}^{\mathbf{C}}\widehat{\mathbf{V}}\right)\right\}\mathbf{1}_N\right\|^2_{\mathbf{R}_t^{-1}},\end{aligned} \tag{66}$$

$$\text{where }\widehat{\mathbf{V}}=\begin{pmatrix}\widehat{\boldsymbol{\chi}}_{(1)}&\cdots&\widehat{\boldsymbol{\chi}}_{(N)}\end{pmatrix}, \tag{67}$$

$$\begin{aligned}\text{and }\delta\mathbf{x}&=\frac{1}{\sqrt{N-1}}\sum_{i=1}^{N}\delta\mathbf{x}^{\mathrm{b}}_{(i)}\circ\left(\mathbf{U}^{\mathbf{C}}\widehat{\boldsymbol{\chi}}_{(i)}\right),\\&=\left\{\mathbf{X}^{\mathrm{b}}\circ\left(\mathbf{U}^{\mathbf{C}}\widehat{\mathbf{V}}\right)\right\}\mathbf{1}_N,\end{aligned} \tag{68}$$

where $\widehat{\boldsymbol{\chi}}_{(i)}$ (n-elements) is the preconditioned version of $\widehat{\boldsymbol{\alpha}}_{(i)}$, $\widehat{\mathbf{V}}$ ($n\times N$) is the preconditioned version of $\widehat{\mathbf{A}}$ and $\mathbf{U}^{\mathbf{C}}$ is an $n\times n$ potentially full-rank square-root of $\mathbf{C}$, the localization matrix in (59). The relationships between the preconditioned and unpreconditioned variables are $\widehat{\boldsymbol{\alpha}}_{(i)}=\mathbf{U}^{\mathbf{C}}\widehat{\boldsymbol{\chi}}_{(i)}$ and $\widehat{\mathbf{A}}=\mathbf{U}^{\mathbf{C}}\widehat{\mathbf{V}}$. The unpreconditioned and preconditioned forms are identical provided that $\mathbf{U}^{\mathbf{C}}\mathbf{U}^{\mathbf{C}\mathrm{T}}=\mathbf{C}$ (as in section 6.2).

In a similar way to the 4D-Var control variable $\delta\boldsymbol{\chi}_{\mathrm{var}}$ in section 2.3, the components of $\widehat{\boldsymbol{\chi}}_{(i)}$ in (66) have unit background covariance within and between control vectors, $\left\langle\widehat{\boldsymbol{\chi}}_{(i)}\widehat{\boldsymbol{\chi}}^{\mathrm{T}}_{(j)}\right\rangle_{\mathrm{b}}=\mathbf{I}\delta_{ij}$, where $\langle\bullet\rangle_{\mathrm{b}}$ indicates expectation over the background distribution. The key to understanding the L03 formulation is given in Appendix E, which shows that the background-error covariance matrix implied by (68) is, as for B05, $\mathbf{P}^{\mathrm{b}}_{(N)}\circ\mathbf{C}$.

There are Nn control vector elements in total in this formulation. Equation (68) is just an extension of (36), where each control vector element in (36) itself becomes a vector. Following Lorenc (2003), we choose to assemble the $\widehat{\boldsymbol{\chi}}_{(i)}$ vectors in a matrix (here $\widehat{\mathbf{V}}$) as in (67). Other authors (e.g. Wang and Lei, 2014; Lorenc *et al.*, 2015) instead choose to assemble these vectors as a long Nn-element vector, but the two representations are equivalent.

6.4. *Comments on localization*

Compact representations B05 and L03 can help to reduce noise in the sample error covariance matrix. However localization (and inflation) are only a partial solution to the under-sampling problem because the moderation functions can modify some important properties of the raw ensemble, e.g. its balance properties (Lorenc, 2003; Houtekamer and Mitchell, 2005; Kepert, 2009; Greybush *et al.*, 2011; Bannister, 2015). This is one justification for developing the hybrid methods (section 5). Hybrid methods still require localization so either B05 or L03 are used in most hybrid systems (section 7 summarises localization in most ensemble-related methods described so far).

In B05 and L03, the specification of the matrix $\mathbf{U}^{\mathbf{C}}$ remains, which can still be a huge matrix. There are ways of modelling $\mathbf{U}^{\mathbf{C}}$ instead of storing large matrices. The approach is identical to that used to model the spatial component of $\mathbf{B}_0$ in Var, e.g. using homogeneous and isotropic functions (Berre, 2000; Bannister, 2008; Errera and Menard, 2012), or more generally using using recursive filters (Hayden and Purser, 1995), wavelets (Fisher, 2003; Deckmyn and Berre, 2005; Bannister, 2007), or spectral localization (Buehner and Charron, 2007). Methods also exist that generate adaptive moderation functions, which depend upon the flow itself (Bishop and Hodyss 2007,2009a,2009b,2011).

The method of Buehner (2012) in particular is a cross between spatial and spectral localization. It is useful in the context of B05 localization, where, instead of (61), it first splits each $\delta\mathbf{x}^{\mathrm{b}}_{(i)}$ into contributions from J wavebands by spectral filtering. For

member i and waveband j this gives $\delta\boldsymbol{\epsilon}^{j}_{(i)} = \mathbf{S}^{-1}\boldsymbol{\Psi}^{j}\mathbf{S}\delta\mathbf{x}^{\mathrm{b}}_{(i)}$, where $\mathbf{S}$ is a spectral transform, and $\boldsymbol{\Psi}^{j}$ is the diagonal bandpass filter associated with band j. Under this scheme, the alternative version of (61) is

$$\widehat{\mathbf{X}}^{\mathrm{b}} = \frac{1}{\sqrt{N-1}}\Big(\mathrm{diag}(\delta\boldsymbol{\epsilon}^{1}_{(1)})\mathbf{U}^{\mathbf{C}1},\ldots,\mathrm{diag}(\delta\boldsymbol{\epsilon}^{1}_{(N)})\mathbf{U}^{\mathbf{C}1},\ldots \mathrm{diag}(\delta\boldsymbol{\epsilon}^{J}_{(1)})\mathbf{U}^{\mathbf{C}J},\ldots,\mathrm{diag}(\delta\boldsymbol{\epsilon}^{J}_{(N)})\mathbf{U}^{\mathbf{C}J}\Big), \quad (69)$$

where $\mathbf{U}^{\mathbf{C}j}$ is a square-root of a localization matrix associated with band j, i.e. $\mathbf{U}^{\mathbf{C}j}\mathbf{U}^{\mathbf{C}j\mathrm{T}} = \mathbf{C}^{j}$. This form allows band (i.e. scale-selective) localization, which is useful because less localization (i.e. larger localization length-scales) is needed for lower wavebands (i.e. larger error scales) than for higher wavebands (Buehner, 2012). This method effectively increases the potential rank of the background-error covariances by a factor of J, which results in the need for fewer ensemble members for similar noise level (Buehner, 2012). This method is extended in (Buehner and Shlyaeva, 2015).

B05 and L03 localizations are equivalent, but their efficiencies depend upon the chosen rank of $\mathbf{U}^{\mathbf{C}}$. Consider for instance an implementation of localization in pure En4DVar.

- The total number of control vector elements required for B05 is NM, making up $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$ in (62). Elements of $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$ are associated with NM perturbation vectors in $\widehat{\mathbf{X}}^{\mathrm{b}}$. For lower-rank implementations of $\mathbf{U}^{\mathbf{C}}$ (where $M \ll n$), B05 is the more efficient approach. B05 is used, for example, in the Environment Canada global and regional deterministic DA systems (Buehner *et al.*, 2015a; Caron *et al.*, 2015), and also in the recent work of Liu and Xue (2016).
- The total number of control vector elements required for L03 is Nn (comprising $\widehat{\mathbf{V}}$ in (66)–(68)), which are associated with the N perturbation vectors in $\mathbf{X}^{\mathrm{b}}$. For full-rank implementations of $\mathbf{U}^{\mathbf{C}}$, L03 is the more efficient approach (since $\widehat{\mathbf{X}}^{\mathrm{b}}$ does not have to be stored) as long as $\mathbf{U}^{\mathbf{C}}$ can be modelled efficiently. L03 is used, for example, in the Met Office's global hybrid system (Clayton *et al.*, 2012).

We note that it is arguably more straightforward to account for model error in 4DEnVar schemes that use B05 localization (section 8).

7. Summary of the EnKF, pure EnVar, hybrid EnVar

We have presented many options for approximating the DA problem which incorporate an ensemble in one way or another. These have been classified as either pure ensemble (EnKF/S, section 3), pure EnVar (essentially Var approaches to finding the mean state of the pure ensemble system, section 4) and hybrid EnVar (extending EnVar with added information from the $\mathbf{B}_0$-matrix of Var, section 5). The B05 (section 6.2) and L03 (section 6.3) localization formulations are now applied to these DA systems to give the final forms of the equations that operational DA can deal with. It is possible (within bounds) to mix-and-match a DA method with a localization method, leaving us with a relatively large number of combinations.

7.1. Localization in the EnKF

The most common way of localizing $\mathbf{P}^{\mathrm{b}}_{(N)}$ in an EnKF (④ in Figure 1) is by restricting the observations that are allowed to affect each analysis point (discussion in section 6). However, to localize in model space, B05 is the appropriate implicit method. As described in section 6.2, a localized version of (28) is found by replacing $\mathbf{P}^{\mathrm{b}} \rightarrow \widehat{\mathbf{P}}^{\mathrm{b}}$ in the first line, which is equivalent to replacing $\mathbf{X}^{\mathrm{b}} \rightarrow \widehat{\mathbf{X}}^{\mathrm{b}}$ in the last line, where $\widehat{\mathbf{X}}^{\mathrm{b}}$ is defined in (61) (not shown).

To avoid linearized operators in the resulting expression, $\mathbf{H}\widehat{\mathbf{X}}^{\mathrm{b}}$ can be approximated by the nonlinear operations $\boldsymbol{\mathcal{H}}(\overline{\mathbf{x}^{\mathrm{b}}}+\delta\widehat{\mathbf{x}}^{\mathrm{b}}_{(i)})-\boldsymbol{\mathcal{H}}(\overline{\mathbf{x}^{\mathrm{b}}})$ (approximating the ith column of $\mathbf{H}\widehat{\mathbf{X}}^{\mathrm{b}}$, where $\delta\widehat{\mathbf{x}}^{\mathrm{b}}_{(i)}$ is the ith column of $\widehat{\mathbf{X}}^{\mathrm{b}}$, and $\overline{\mathbf{x}^{\mathrm{b}}}$ is the ensemble mean).

7.2. Model space localization in the EnKS

7.2.1. Localization in the EnKS (3D state vector)

The EnKS with a 3D state vector (where observations are predicted by combining the forecast model with the observation operator (29), ⑤ in Figure 1) may also be localized with B05; in fact it would be difficult to use the observation space localization mentioned in section 6 with this EnKS, as the presence of the forecast model means that even point observations for $t > 0$ cannot be associated with points in the model domain at $t = 0$. This localized EnKS is found by a identical procedure to that described in section 7.1, applied to (29) (not shown).

To avoid linearized operators in the resulting expression, $\underline{\mathbf{H}}_{\mathbf{M}}\widehat{\mathbf{X}}^{\mathrm{b}}$ can be approximated by $\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\overline{\mathbf{x}^{\mathrm{b}}}+\delta\widehat{\mathbf{x}}^{\mathrm{b}}_{(i)})-\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\overline{\mathbf{x}^{\mathrm{b}}})$ (approximating the ith column of $\underline{\mathbf{H}}_{\mathbf{M}}\widehat{\mathbf{X}}^{\mathrm{b}}$).

7.2.2. Localization in the EnKS (4D state vector)

The EnKS with a 4D state vector (⑥ in Figure 1) uses $\underline{\mathbf{X}}^{\mathrm{b}}$ instead of $\mathbf{X}^{\mathrm{b}}$ (32), which eliminates the forecast model from within the DA (33). The localization must also be extended to the time dimension, $\mathbf{C} \rightarrow \underline{\mathbf{C}}$, so $\mathbf{U}^{\mathbf{C}} \rightarrow \underline{\mathbf{U}}^{\mathbf{C}}$, where $\underline{\mathbf{C}} = \underline{\mathbf{U}}^{\mathbf{C}}\underline{\mathbf{U}}^{\mathbf{C}\mathrm{T}}$. This version of the localized EnKS is found by a similar procedure to that described in sections 7.1 and 7.2.1, applied to (33), by replacing $\underline{\mathbf{X}}^{\mathrm{b}} \rightarrow \widehat{\underline{\mathbf{X}}}^{\mathrm{b}}$ (not shown).

As above, each column of $\underline{\mathbf{H}}\widehat{\underline{\mathbf{X}}}^{\mathrm{b}}$ can be approximated by the nonlinear operations $\underline{\boldsymbol{\mathcal{H}}}(\overline{\underline{\mathbf{x}}^{\mathrm{b}}}+\delta\widehat{\underline{\mathbf{x}}}^{\mathrm{b}}_{(i)})-\underline{\boldsymbol{\mathcal{H}}}(\overline{\underline{\mathbf{x}}^{\mathrm{b}}})$ (approximating the ith column of $\underline{\mathbf{H}}\widehat{\underline{\mathbf{X}}}^{\mathrm{b}}$).

7.2.3. Comments on the implied localized background-error covariances in these systems

Due to the way that the different localizations work in the above 3D and 4D versions of the EnKS, they do not necessarily share the same implied localized background-error covariances, even when the forecast model is linear. In order to examine how they differ, consider the implied error covariance matrix between fields at time t. This is studied by looking at the analysis increment (propagated to time t) due to observations at (only) t. Under these conditions the analysis increment of the 3D state vector form of the EnKS (29) ($\mathbf{X}^{\mathrm{b}} \rightarrow \widehat{\mathbf{X}}^{\mathrm{b}}$) is:

$$\begin{aligned}
&\mathbf{M}_{0,t}(\mathbf{x}^{\mathrm{a}}_{(i)}-\mathbf{x}^{\mathrm{b}}_{(i)})\\
&=\mathbf{M}_{0,t}\widehat{\mathbf{X}}^{\mathrm{b}}\left(\mathbf{H}_t\mathbf{M}_{0,t}\widehat{\mathbf{X}}^{\mathrm{b}}\right)^{\mathrm{T}}\\
&\quad\times\left\{\mathbf{R}_t+\mathbf{H}_t\mathbf{M}_{0,t}\widehat{\mathbf{X}}^{\mathrm{b}}\left(\mathbf{H}_t\mathbf{M}_{0,t}\widehat{\mathbf{X}}^{\mathrm{b}}\right)^{\mathrm{T}}\right\}^{-1}\\
&\quad\times\left\{\mathbf{y}^{\mathrm{o}}_{(i)t}-\boldsymbol{\mathcal{H}}_t\left[\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(i)})\right]\right\}\\
&=\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t)\mathbf{H}_t^{\mathrm{T}}\left\{\mathbf{R}_t+\mathbf{H}_t\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t)\mathbf{H}_t^{\mathrm{T}}\right\}^{-1}\\
&\quad\times\left\{\mathbf{y}^{\mathrm{o}}_{(i)t}-\boldsymbol{\mathcal{H}}_t\left[\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}}_{(i)})\right]\right\},
\end{aligned} \quad (70)$$

where $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t) = \mathbf{M}_{0,t}\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}\mathbf{M}^{\mathrm{T}}_{0,t} = \mathbf{M}_{0,t}\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}\mathbf{M}^{\mathrm{T}}_{0,t}$, and where $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)} = \widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(0) = \mathbf{C}\circ\mathbf{P}^{\mathrm{b}}_{(N)}$. As usual, the hat (no hat) means 'localized' ('unlocalized'). This analysis allows us to see that $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t)$ is the propagated localized background-error covariance implied in the localized EnKS with 3D state vector, which is just the localized covariance at $t = 0$ propagated with the linear model.

The analysis increment of the 4D state vector form of the EnKS (33) ($\underline{\mathbf{X}}^{\mathrm{b}} \rightarrow \underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$) due to observations at (only) t is now studied. It exists at all times in the window, but the increment at t can be isolated with the operator comprising $T+1$ blocks, each $n \times n$: $\mathbf{I}_t = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{I} & \cdots & \mathbf{0} \end{pmatrix}$, where all blocks are $\mathbf{0}$, except for the one corresponding to time t, which is $\mathbf{I}$. Noting that for this case $\underline{\mathbf{H}}\underline{\widehat{\mathbf{X}}}^{\mathrm{b}} = \mathbf{H}_t\widehat{\mathbf{X}}^{\mathrm{b}}(t)$, and $\mathbf{I}_t\underline{\widehat{\mathbf{X}}}^{\mathrm{b}} = \widehat{\mathbf{X}}^{\mathrm{b}}(t)$, the time t analysis increment from (33) ($\underline{\mathbf{X}}^{\mathrm{b}} \rightarrow \underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$) is

$$\begin{aligned}\mathbf{x}^{\mathrm{a}}_{(i)}(t) - \mathbf{x}^{\mathrm{b}}_{(i)}(t) &= \mathbf{I}_t\left(\underline{\mathbf{x}}^{\mathrm{a}}_{(i)} - \underline{\mathbf{x}}^{\mathrm{b}}_{(i)}\right) \\ &= \widehat{\mathbf{X}}^{\mathrm{b}}(t)\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}(t)\mathbf{H}_t^{\mathrm{T}}\left\{\mathbf{R}_t + \mathbf{H}_t\widehat{\mathbf{X}}^{\mathrm{b}}(t)\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}(t)\mathbf{H}_t^{\mathrm{T}}\right\}^{-1} \\ &\quad \times \left\{\mathbf{y}^{\mathrm{o}}_{(i)}(t) - \boldsymbol{\mathcal{H}}_t[\mathbf{x}^{\mathrm{b}}_{(i)}(t)]\right\}. \end{aligned} \tag{71}$$

This analysis allows us to see that the localized covariance matrix implied here at time t is $\widehat{\mathbf{X}}^{\mathrm{b}}(t)\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}(t) = \mathbf{I}_t\left(\mathbf{C}\circ\underline{\mathbf{P}}^{\mathrm{b}}_{(N)}\right)\mathbf{I}_t^{\mathrm{T}} = \mathbf{C}(t)\circ\mathbf{P}^{\mathrm{b}}_{(N)}(t)$. The differences between this covariance and the one for the 3D state vector are that (i) here no linearized model is used and (ii) the localization at time t, $\mathbf{C}(t)$ here needs to be specified (via $\underline{\mathbf{U}}^{\mathbf{C}}$), rather than being propagated from $\mathbf{C}(0)$. This difference also separates the localized En4DVar and 4DEnVar schemes (below).

7.3. Localization and hybridization in En4DVar

In this section, En4DVar (the method in section 4.1 of finding the mean of the EnKS variationally, making use of the linear models of 4D-Var) is localized with B05 and with L03 (⓫ in Figure 1, and ⓫ is for hybrid En3DVar). The equations shown are also hybridized with the climatological $\mathbf{B}_0$.

7.3.1. B05 localization in hybrid En4DVar

The En4DVar equations hybridized and localized with B05 emerge from a straightforward application of concepts in section 5.3 and 6.2 to (35). The cost function, CVT and gradient formulae for this case are given as:

$$\widehat{J}^{\mathrm{HEn4DVar}}(\delta\boldsymbol{\chi}_{\mathrm{var}}, \widehat{\boldsymbol{\chi}}_{\mathrm{ens}}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\delta\mathbf{x}\right\|^2_{\underline{\mathbf{R}}^{-1}}, \tag{72}$$

$$\delta\mathbf{x} = \sqrt{1-\beta}\mathbf{U}\delta\boldsymbol{\chi}_{\mathrm{var}} + \sqrt{\beta}\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}, \tag{73}$$

$$\nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{h}}}\widehat{J}^{\mathrm{HEn4DVar}} = \begin{pmatrix} \nabla_{\boldsymbol{\chi}_{\mathrm{var}}}\widehat{J}^{\mathrm{HEn4DVar}} \\ \nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}}\widehat{J}^{\mathrm{HEn4DVar}} \end{pmatrix} = \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} - \sqrt{1-\beta}\mathbf{U}^{\mathrm{T}}\underline{\mathbf{H}}_{\mathbf{M}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\delta\mathbf{x}\right) \\ \widehat{\boldsymbol{\chi}}_{\mathrm{ens}} - \sqrt{\beta}\widehat{\mathbf{X}}^{\mathrm{b}\mathrm{T}}\underline{\mathbf{H}}_{\mathbf{M}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\delta\mathbf{x}\right) \end{pmatrix}, \tag{74}$$

where $\underline{\mathbf{H}}_{\mathbf{M}}$ is defined in (31), $\underline{\mathbf{d}}$ is defined after (33), and we define the augmented (hybridized) control vector as $\widehat{\boldsymbol{\chi}}_{\mathrm{h}} = \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} & \widehat{\boldsymbol{\chi}}_{\mathrm{ens}} \end{pmatrix}$, which has $n+NM$ elements in total. Here $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$ is the NM-element vector associated with the N ensemble members in $\mathbf{X}^{\mathrm{b}}$ combined with the M localization members in $\mathbf{U}^{\mathbf{C}}$ (61). The differences with the original form (35) due to localization (75) include the extended length of the control vector ($\boldsymbol{\chi}_{\mathrm{ens}} \rightarrow \widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$), and $\mathbf{X}^{\mathrm{b}} \rightarrow \widehat{\mathbf{X}}^{\mathrm{b}}$.

7.3.2. L03 localization in hybrid En4DVar

The cost function for En4DVar with L03 localization (66) was used as the example in section 6.3. Incorporating the hybrid scheme from section 5.3 leads to the cost function, CVT and gradient formulae for this case given as:

$$\widehat{J}^{\mathrm{HEn4DVar}}(\delta\boldsymbol{\chi}_{\mathrm{var}}, \widehat{\mathbf{V}}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\sum_{i=1}^{N}\left\|\widehat{\boldsymbol{\chi}}_{(i)}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\delta\mathbf{x}\right\|^2_{\underline{\mathbf{R}}^{-1}}, \tag{75}$$

$$\delta\mathbf{x} = \sqrt{1-\beta}\mathbf{U}\delta\boldsymbol{\chi}_{\mathrm{var}} + \sqrt{\beta}\left\{\mathbf{X}^{\mathrm{b}}\circ(\mathbf{U}^{\mathbf{C}}\widehat{\mathbf{V}})\right\}\mathbf{1}_N, \tag{76}$$

$$\nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{h}}}\widehat{J}^{\mathrm{HEn4DVar}} = \begin{pmatrix} \nabla_{\boldsymbol{\chi}_{\mathrm{var}}}\widehat{J}^{\mathrm{HEn4DVar}} \\ \nabla_{\widehat{\mathbf{V}}}\widehat{J}^{\mathrm{HEn4DVar}} \end{pmatrix} = \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} - \sqrt{1-\beta}\mathbf{U}^{\mathrm{T}}\underline{\mathbf{H}}_{\mathbf{M}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{H}}_{\mathbf{M}}\delta\mathbf{x}\right) \\ \widehat{\mathbf{V}} + \sqrt{\beta}\mathbf{U}^{\mathbf{C}\mathrm{T}}\mathbf{X}^{\mathrm{b}}\circ\begin{pmatrix} \nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}} & \cdots & \nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}} \end{pmatrix} \end{pmatrix}, \tag{77}$$

where we define the augmented (hybridized) control vector for L03 as $\widehat{\boldsymbol{\chi}}_{\mathrm{h}} = \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} & \widehat{\mathbf{V}} \end{pmatrix}$, which has $n+Nn$ elements in total. Note that the ensemble part, $\widehat{\mathbf{V}}$, comprises N n-element vectors ($\widehat{\boldsymbol{\chi}}_{(i)}$, assembled here as columns of matrix $\widehat{\mathbf{V}}$, so it should strictly be called a control *matrix*). The gradient of (75) with respect to $\widehat{\mathbf{V}}$ (forming the $n\times N$ sub-matrix) gives the lower part of (77), which has matrix elements

$$\left[\nabla_{\widehat{\mathbf{V}}}\widehat{J}^{\mathrm{HEn4DVar}}\right]_{ij} = \partial\widehat{J}^{\mathrm{HEn4DVar}}/\partial(\widehat{\boldsymbol{\chi}}_{(i)})_j.$$

Appendix B derives this part of the gradient, where $\begin{pmatrix} \nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}} & \cdots & \nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}} \end{pmatrix}$ is the $n\times N$ matrix of repeated columns of $\nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}}$, which is the gradient of the observation term of the strong-constraint 4D-Var cost function (15). The 4D nature of the problem is wrapped up in $\nabla_{\delta\mathbf{x}}J^{4\mathrm{DVar}}_{\mathrm{o}}$.

This hybrid En4DVar system is essentially that of Clayton *et al.* (2012), but with a number of differences in the detail. Clayton *et al.* (2012) have an additional term in the cost function, J_{c} (a term to penalize imbalance). Another is that Clayton *et al.* (2012) perform localization in the space of 'parameter perturbations' instead of model variable perturbations. These parameters are related to the model perturbations via relationships that include balance relations. This is a way (in addition to J_{c}) of introducing balance into the analysis increments which would have been disrupted if localization was done directly in the space of model variables.

7.4. Localization and hybridization in 4DEnVar

In this section, 4DEnVar (the method in section 4.2 of finding the mean of the EnKS variationally, without the need for the linear models) is localized with B05 and with L03 (⓭ in Figure 1). As for En4DVar, the equations shown are also hybridized with $\mathbf{B}_0$.

7.4.1. B05 localization in hybrid 4DEnVar

The 4DEnVar equations hybridized and localized with B05 emerge from an application of concepts in sections 5.3 and 6.2 to (40). The cost function, CVT and gradient formulae for this case are given as:

$$\widehat{J}^{\mathrm{H4DEnVar}}(\delta\boldsymbol{\chi}_{\mathrm{var}}, \widehat{\boldsymbol{\chi}}_{\mathrm{ens}}) = \frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}\right\|^2_{\mathbf{I}} + \frac{1}{2}\left\|\underline{\mathbf{d}} - \underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right\|^2_{\underline{\mathbf{R}}^{-1}}, \tag{78}$$

$$\delta\underline{\mathbf{x}} = \sqrt{1-\beta}\underline{\mathbf{U}}\delta\boldsymbol{\chi}_{\mathrm{var}} + \sqrt{\beta}\underline{\widehat{\mathbf{X}}}^{\mathrm{b}}\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}, \tag{79}$$

$$\nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{h}}}\widehat{J}^{\mathrm{H4DEnVar}} = \begin{pmatrix} \nabla_{\boldsymbol{\chi}_{\mathrm{var}}}\widehat{J}^{\mathrm{H4DEnVar}} \\ \nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}}\widehat{J}^{\mathrm{H4DEnVar}} \end{pmatrix} = \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} - \sqrt{1-\beta}\underline{\mathbf{U}}^{\mathrm{T}}\underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right) \\ \widehat{\boldsymbol{\chi}}_{\mathrm{ens}} - \sqrt{\beta}\underline{\widehat{\mathbf{X}}}^{\mathrm{b}\mathrm{T}}\underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}} - \underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right) \end{pmatrix}, \tag{80}$$

where $\underline{\mathbf{H}}$ is defined in (31), $\underline{\mathbf{d}}$ is defined after (33), and the hybridized control vector has the same structure as in section 7.3.1.

Apart from the 4D nature of the ensemble-related matrices like $\underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$, notice that $\underline{\mathbf{U}}$ (now underlined) now needs to be a 4D operator to make up for the lack of the linearized forecast model to propagate perturbations from $t=0$. This is an important difference between hybrid En4DVar and hybrid 4DEnVar. The are possible ways to structure $\underline{\mathbf{U}}$ and some are listed below.

- Let $\underline{\mathbf{U}}$ be a persistence model of climatological perturbations (Desroziers *et al.*, 2014; Lorenc *et al.*, 2015):

$$\underline{\mathbf{U}}=\begin{pmatrix}\mathbf{I}\\ \vdots\\ \mathbf{I}\end{pmatrix}\mathbf{U}, \tag{81}$$

  where $\mathbf{U}$ is the $t=0$ CVT as used in (17) in section 2.3. The preceding matrix contains $T+1$ blocks, each comprising the $n\times n$ identity matrix. The contribution from $\sqrt{1-\beta}\underline{\mathbf{U}}\delta\boldsymbol{\chi}_{\mathrm{var}}$ in (79) then behaves as 3D-FGAT (section 2.1.3 and Poterjoy and Zhang, 2015) and the climatological component of the implied background-error covariance between any two state-vector components is independent of time. Even though this is classified as a 4D method, it does not exploit the climatological covariance propagation property of 4D-Var. This is the simplest option which is used with most, if not all, applications of hybrid 4DEnVar at present.
- A more sophisticated form of $\underline{\mathbf{U}}$ is to extend a 3D covariance model used to represent $\mathbf{U}$ to 4D. A strategy for modelling 3D background-error covariances is to exploit the normal modes of the system (e.g. Žagar *et al.*, 2004), where the control vector represents coefficients that weight perturbations as linear combinations of 3D normal modes. This could in principle be extended to 4D (where columns of $\underline{\mathbf{U}}$ would be related to the evolving normal modes of the system) although would be limited by linearity of the forecast model.
- A further option is to replace the climatological term of 4DEnVar with that of En4DVar, but use a highly simplified model – e.g. with advection only, $\mathbf{M}^{\mathrm{adv}}_{0,t}$. The computational cost issues associated with use of linear models (and adjoints) would remain, but such a simple model would require little maintenance.

It might also be possible to propagate the localization square-root matrix in 4DEnVar (either with the full, linearized or simplified model). Column i of the square-root matrix $\underline{\mathbf{U}}^{\mathbf{C}}$ (section 6.2, call $\underline{\mathbf{u}}^{\mathbf{C}}_{(i)}$) would then, e.g. comprise the pre-computed column vector $\underline{\mathbf{u}}^{\mathbf{C}}_{(i)}=\left(\mathbf{u}^{\mathbf{C}}_{(i)}(0),\mathbf{M}^{\mathrm{adv}}_{0,1}\mathbf{u}^{\mathbf{C}}_{(i)}(0),\ldots,\mathbf{M}^{\mathrm{adv}}_{T-1,T}\mathbf{u}^{\mathbf{C}}_{(i)}(T-1)\right)$. This, and ideas of P. Arbogast (2016; personal communication) are possible ways of dealing with the 4D aspects of localization.

7.4.2. *L03 localization in hybrid 4DEnVar*

The hybrid 4DEnVar equations with L03 localization emerge from an application of concepts in sections 5.3 and 6.3 to (40). The cost function, CVT and gradient formulae for this case are given as:

$$\widehat{J}^{\mathrm{H4DEnVar}}(\delta\boldsymbol{\chi}_{\mathrm{var}},\underline{\widehat{\mathbf{V}}})=\frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|^2_{\mathbf{I}}+\frac{1}{2}\sum_{i=1}^{N}\left\|\widehat{\boldsymbol{\chi}}_{(i)}\right\|^2_{\mathbf{I}}+\frac{1}{2}\left\|\underline{\mathbf{d}}-\underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right\|^2_{\underline{\mathbf{R}}^{-1}}, \tag{82}$$

$$\delta\underline{\mathbf{x}}=\sqrt{1-\beta}\underline{\mathbf{U}}\delta\boldsymbol{\chi}_{\mathrm{var}}+\sqrt{\beta}\left\{\underline{\mathbf{X}}^{\mathrm{b}}\circ(\underline{\mathbf{U}}^{\mathbf{C}}\underline{\widehat{\mathbf{V}}})\right\}\mathbf{1}_N, \tag{83}$$

$$\nabla_{\widehat{\boldsymbol{\chi}}_{\mathrm{h}}}\widehat{J}^{\mathrm{H4DEnVar}}=\begin{pmatrix}\nabla_{\boldsymbol{\chi}_{\mathrm{var}}}\widehat{J}^{\mathrm{H4DEnVar}}\\ \nabla_{\underline{\widehat{\mathbf{V}}}}\widehat{J}^{\mathrm{H4DEnVar}}\end{pmatrix}$$

$$=\begin{pmatrix}\delta\boldsymbol{\chi}_{\mathrm{var}}-\sqrt{1-\beta}\underline{\mathbf{U}}^{\mathrm{T}}\underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{d}}-\underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right)\\ \underline{\widehat{\mathbf{V}}}+\sqrt{\beta}\underline{\mathbf{U}}^{\mathbf{C}\mathrm{T}}\underline{\mathbf{X}}^{\mathrm{b}}\circ\left(\nabla_{\delta\underline{\mathbf{x}}}J^{4\mathrm{DVar}}_{\mathrm{o}}\ \cdots\ \nabla_{\delta\underline{\mathbf{x}}}J^{4\mathrm{DVar}}_{\mathrm{o}}\right)\end{pmatrix}, \tag{84}$$

where the total control 'vector' $\widehat{\boldsymbol{\chi}}_{\mathrm{h}}=(\delta\boldsymbol{\chi}_{\mathrm{var}},\underline{\mathbf{C}})$ has $n+n(T+1)N$ elements in total. In these forms, virtually all objects are underlined, indicating that they are 4D. This is the way that localized hybrid 4DEnVar avoids the Jacobian of the forecast model (i.e. the forecast part of the problem is dealt with in the preparation of $\underline{\mathbf{X}}^{\mathrm{b}}$). The gradient $\nabla_{\delta\underline{\mathbf{x}}}J^{4\mathrm{DVar}}_{\mathrm{o}}$ that appears (as N repeated columns) in (84) is not found from (15); that equation gives $\nabla_{\delta\mathbf{x}(0)}$, the gradient with respect to the initial perturbation in 4D-Var. Instead $\nabla_{\delta\underline{\mathbf{x}}}J^{4\mathrm{DVar}}_{\mathrm{o}}$ is the following $n(T+1)$-element vector that does not require the model adjoint:

$$\nabla_{\delta\underline{\mathbf{x}}}J^{4\mathrm{DVar}}_{\mathrm{o}}=\begin{pmatrix}\nabla_{\delta\mathbf{x}(0)}J^{4\mathrm{DVar}}_{\mathrm{o}}\\ \vdots\\ \nabla_{\delta\mathbf{x}(T)}J^{4\mathrm{DVar}}_{\mathrm{o}}\end{pmatrix}=\begin{pmatrix}-\mathbf{H}^{\mathrm{T}}_0\mathbf{R}^{-1}_0\left\{\mathbf{d}_0-\mathbf{H}_0\delta\mathbf{x}(0)\right\}\\ \vdots\\ -\mathbf{H}^{\mathrm{T}}_T\mathbf{R}^{-1}_T\left\{\mathbf{d}_T-\mathbf{H}_T\delta\mathbf{x}(T)\right\}\end{pmatrix}, \tag{85}$$

where $\mathbf{d}_t$ is defined by (12).

Note that $\underline{\widehat{\mathbf{V}}}$ is a much larger matrix than $\widehat{\mathbf{V}}$ is in hybrid En4DVar with L03 localization. The other difficulty over En4DVar is that a 4D localization must be modelled via $\underline{\mathbf{U}}^{\mathbf{C}}$ (discussion in section 7.2.3). This, and the '$\underline{\mathbf{U}}$ problem' are important outstanding issues in the development of hybrid 4DEnVar (the discussion in section 7.4.1 still stands for L03 localization).

7.4.3. *Comments on the Jacobian of the observation operator and on the implied localized background-error covariances in these systems*

Hybrid 4DEnVar in (78) and (82) retains the Jacobian $\underline{\mathbf{H}}$. This could be avoided whenever $\underline{\mathbf{H}}\underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$ appears where it can be approximated with differences between runs of the nonlinear observation operator; e.g. for perturbations at time t at column i of $\underline{\widehat{\mathbf{X}}}^{\mathrm{b}}$ the approximation would be

$$\boldsymbol{\mathcal{H}}_t\left\{\mathbf{x}^{\mathrm{b}}(t)+\delta\widehat{\mathbf{x}}^{\mathrm{b}}_{(i)}(t)\right\}-\boldsymbol{\mathcal{H}}_t\left\{\mathbf{x}^{\mathrm{b}}(t)\right\}.$$

There is no obvious way of avoiding $\underline{\mathbf{H}}$ elsewhere.

The ensemble part of the implied localized background-error covariances between localized En4DVar and localized 4DEnVar differ in the same way as those between the 3D and 4D versions of the EnKS (section 7.2.3). Having a 3D state vector (where future states are generated by the model within the DA), En4DVar has the implied covariance at time t: $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t)=\mathbf{M}_{0,t}\left(\mathbf{C}\circ\mathbf{P}^{\mathrm{b}}_{(N)}\right)\mathbf{M}^{\mathrm{T}}_{0,t}$ (quantities in the brackets are valid at $t=0$), and having a 4D state vector, 4DEnVar has the implied covariance at time t: $\widehat{\mathbf{P}}^{\mathrm{b}}_{(N)}(t)=\mathbf{C}(t)\circ\mathbf{P}^{\mathrm{b}}_{(N)}(t)$. This result is independent of whether

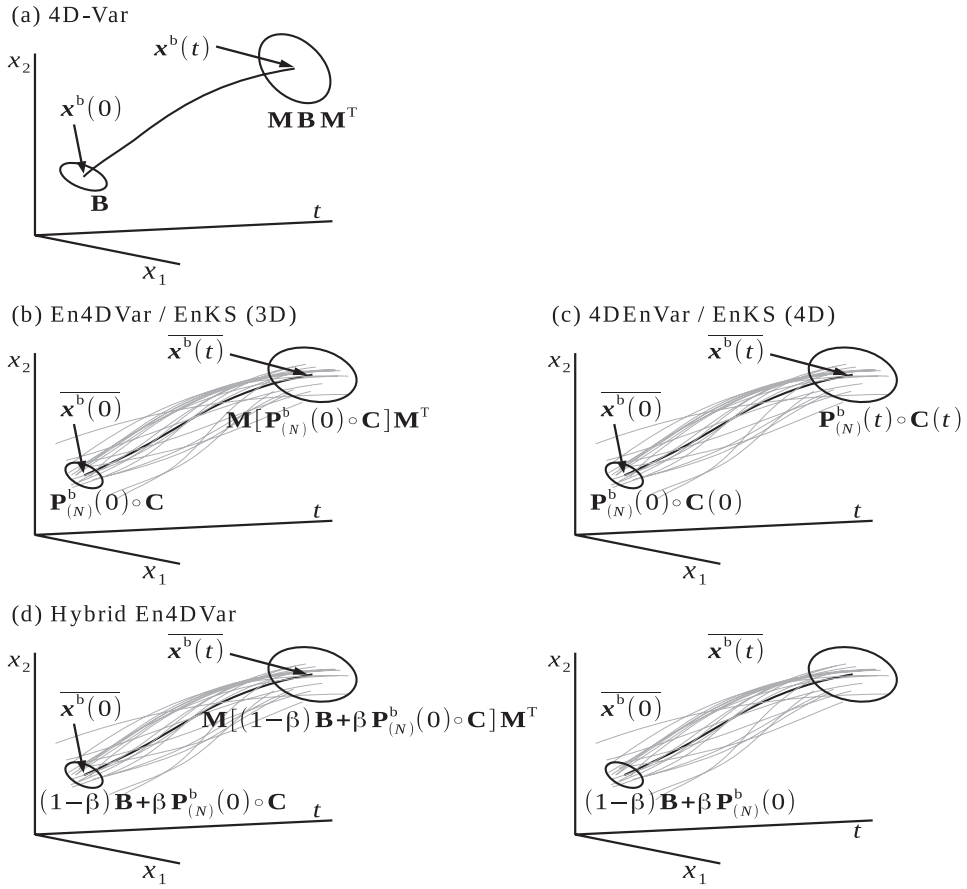

Figure 2. How the background-error covariance matrix evolves throughout the assimilation window for various 4D methods. (a) 4D-Var, (b) En4DVar (or the EnKS with 3D state vector), (c) 4DEnVar (or the EnKS with 4D state vector), (d) hybrid En4DVar and (e) hybrid 4DEnVar. For simplicity, model error is not considered, and $\mathbf{M}$ is short for $\mathbf{M}_{0,t}$. In hybrid 4DEnVar (e), $\mathbf{B}$ is not evolved (as is current practice). The ellipses represent the background-error covariance 'bubbles' at times 0 and t, and the curves represent background ensemble members, where the thick black curve in each panel is the ensemble mean. In all Var techniques, only a single analysis trajectory is conventionally found.

the B05 or the L03 formulation is used. Figure 2 gives a pictorial representation of how the background-error covariances evolve throughout the assimilation window in each 4D scheme discussed in this article.

7.5. *Summary of nomenclature*

Figure 1 serves to summarize the nomenclature, including the names recommended by (Lorenc, 2013) and in this article, and alternative names. It also summarizes the basic features of the localized schemes including the number of control variables (the Figure caption gives more information). As is the case throughout this article, we recognise, but do not elaborate on, the spectrum of methods that count as an EnKF (e.g. ETKF, etc.).

8. Model error with (hybrid) En4DVar and 4DEnVar

No assimilation is optimum until it accounts correctly for all sources of error. Model error is an important source, which is a particular concern when the assimilation window is 'long'. We have already considered model error with weak-constraint 4D-Var (section 2.1), and here we consider how model error might be represented in En4DVar and 4DEnVar.

8.1. *Model error in (hybrid) En4DVar*

Accounting for model error in methods that use the linearized model, e.g. hybrid En4DVar, may be achieved by augmenting the control variable with the variables associated with model error, $\delta\boldsymbol{\eta}(1),\ldots,\delta\boldsymbol{\eta}(T)$, adding a model error term as in (10), and modifying the prediction operator of the observations at time t. For instance hybrid En4DVar with B05 localization (14 in Figure 1) has the modified model observations (based on (8)):

$$\begin{aligned}\mathbf{y}^{\mathrm{x}}_t =& \boldsymbol{\mathcal{H}}_t\{\boldsymbol{\mathcal{M}}_{0,t}(\mathbf{x}^{\mathrm{b}})\}\\ &+\mathbf{H}_t\Big[\mathbf{M}_{0,t}\left(\sqrt{1-\beta}\mathbf{U}\delta\boldsymbol{\chi}_{\mathrm{var}}+\sqrt{\beta}\widehat{\mathbf{X}}^{\mathrm{b}}\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}\right)\\ &+\sum_{\tau=1}^{t}\mathbf{M}_{\tau,t}\delta\boldsymbol{\eta}(\tau)\Big].\end{aligned} \tag{86}$$

The new terms accumulate the estimated model error contributions up to and including time t. A similar procedure is valid for L03 localization.**

8.2. *Model error in 4DEnVar*

Accounting for model error in methods that do not use the linearized model, like hybrid 4DEnVar, is more in line with the requirements of many operational centres, but is arguably harder to do and there are few examples in the literature. The key is to build flexibility into a scheme to relax the need for the solution to be synthesized from model trajectories over the window. One possibility, based around the B05 localization is to allow a different linear combination of members at each time (Amezcua *et al.*, 2017; Goodliff *et al.*, 2017). Instead of one control vector associated with the ensemble, $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}$ in (79), there would be $T+1$

**In practice the preconditioned versions of the model error variables would be used ((18) of section 2.3).

such vectors. The versions of (78) and (79) become (87) and (88) respectively,

$$\begin{aligned}
&\widehat{J}^{\mathrm{H4DEnVar}}\left\{\delta\boldsymbol{\chi}_{\mathrm{var}},\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(0),\ldots,\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(T)\right\}\\
&=\frac{1}{2}\left\|\delta\boldsymbol{\chi}_{\mathrm{var}}\right\|_{\mathbf{I}}^{2}\\
&+\frac{1}{2}\sum_{t=0}^{T}\left\|\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(t)\right\|_{\mathbf{I}}^{2}+\left\|\underline{\mathbf{y}}-\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\mathrm{b}})-\underline{\mathbf{H}}\delta\underline{\mathbf{x}}\right\|_{\underline{\mathbf{R}}^{-1}}^{2}\\
&+\frac{1}{2}\left\|\widehat{\mathbf{X}}^{\mathrm{b}}(t)\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(t)-\mathbf{M}_{t-1,t}\widehat{\mathbf{X}}^{\mathrm{b}}(t-1)\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(t-1)\right\|_{\mathbf{Q}_{t}^{-1}}^{2},
\end{aligned}\tag{87}$$

$$\delta\underline{\mathbf{x}}=\sqrt{1-\beta}\,\underline{\mathbf{U}}\delta\boldsymbol{\chi}_{\mathrm{var}}+\sqrt{\beta}\begin{pmatrix}\widehat{\mathbf{X}}^{\mathrm{b}}(0)\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(0)\\ \vdots\\ \widehat{\mathbf{X}}^{\mathrm{b}}(T)\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(T)\end{pmatrix},\tag{88}$$

where there are $n+(T+1)NM$ elements to the control vector in total (or $(T+1)NM$ for the non-hybrid version). The last term of (87) penalizes model trajectory misfits, which have covariance $\mathbf{Q}_t$.

There is an inconsistency in system (87) and (88), namely that model error has been accounted for in the part associated with $\widehat{\boldsymbol{\chi}}_{\mathrm{ens}}(t)$, but not in the part associated with $\delta\boldsymbol{\chi}_{\mathrm{var}}$ (the latter is akin to 3DFGAT). Clearly accounting for model error effectively in hybrid 4DEnVar will require a great deal of research.

Another point of interest is the nature of the 4D ensemble comprising $\mathbf{X}^{\mathrm{b}}$ (and hence the 'localized' version $\widehat{\mathbf{X}}^{\mathrm{b}}$). Changing only the linear combination of members at each time step does not allow significant deviations from the 'model attractor', which would be required when the model is in error. An obvious way of dealing with this is to modify the way that the ensemble is prepared by allowing the trajectories to deviate stochastically. There are many well-documented ways of doing this (Palmer and Williams, 2010), e.g. using multiple physical parametrization schemes (Stensrud *et al.*, 2000; Berner *et al.*, 2011), using the stochastic kinetic energy backscatter scheme (Shutts, 2005), perturbing physics tendencies (Buizza *et al.*, 1999; Bouttier *et al.*, 2012; Fresnay *et al.*, 2012), or perturbing physics parameters (Bowler *et al.*, 2008; Hacker *et al.*, 2011; Gebhardt *et al.*, 2011; Vié *et al.*, 2012; Baker *et al.*, 2014). Timing errors may be represented by adding ensemble members with a slightly different validity time to the background time (Gustafsson *et al.*, 2014). Including additive model error during model integration has been done by many authors in the context of the EnKF to good effect (e.g. Houtekamer *et al.*, 2005; Hamill and Whitaker, 2005; Kalnay *et al.*, 2007).

9. Which method to use, and how do the methods compare in use?

There is now a variety of known DA methods, so here we compare the merits and the relative performance of each as reported in the literature.

9.1. Factors to be considered

The decision on which DA method to use depends upon many factors, some of which are listed here.

- Parallelizability: Methods that do not require time integration within the DA algorithm have the edge on parallelizability, since the forward and adjoint integrations need to be done serially. Suitable methods include the EnKS (4D state vector), 4DEnVar, and (hybrid) 4DEnVar. Although the EnKS (3D state vector) and (hybrid) En4DVar do time integration within the DA algorithm this can be allocated to separate processors.
- Existing 4D-Var infrastructure. If linearized and adjoint versions of the forecast model are already available (e.g. used already in 4D-Var), then upgrading to (hybrid) En4DVar may well be beneficial (section 9.2).
- Systems with highly nonlinear processes or frequently upgraded forecast models. It is well known that highly nonlinear models (e.g. with moist processes, switches, etc.) present difficulties with linearizing. There are also complications over linearizing models that have, for example, semi-Lagrangian dynamical cores. Anyway most models are time consuming to linearize and are expensive to run, so methods that avoid the need for linearized and adjoint models are desirable. Methods like the EnKS (4D state vector) and (hybrid) 4DEnVar are desirable in this case.
- Known background-error statistics. In systems that have background-error covariance statistics that are easily modelled and whose variability is negligible then 4D-Var (strong- and weak-constraint) remains useful. This is not normally the case for weather forecasting, even though 4D-Var is historically successful.
- Model error is important. If the model error accumulated over the required assimilation time window is significant compared to other sources of error that are accounted for already, then a method with model error should be considered. Of the methods shown in Figure 1, weak-constraint 4D-Var (section 2.1.2), weak-constraint (hybrid) EnVar (section 8), and the EnKF account for model error. There are other possibilities, e.g. methods that use a perfect model, but absorb model error into the $\mathbf{R}_t$-matrix (Howes, 2016).
- Deterministic or probabilistic. Probabilistic forecasts are increasingly in demand (e.g. Palmer, 2012). If the information contained in a single forecast is not adequate, and measures of analysis and forecast uncertainty are required, then the EnKF, EnKS (and other flavours) are useful. Of course (hybrid) En4DVar and 4DEnVar need an ensemble system running alongside them anyway. It is possible to run these methods many times, by perturbing inputs like the background state and the observations (Liu and Xiao, 2013) or use a scheme like EVIL (section 5.5), which does not need a separate ensemble.
- Non-instantaneous or 'rate-of-change' observations. If observations are to be assimilated that are a function of the state at multiple times (e.g. precipitation accumulation) then methods that have 4D state vectors are the natural choice, e.g. 4D-Var, EnKS and (hybrid) En4DVar/4DEnVar. The equations describing these systems will need to be modified to allow for such observations.

9.2. Systems studied in literature and their relative performance

There are now a large number of studies testing and comparing the DA methods discussed in this article. In this section we summarize the main findings. We consider three kinds of study:
(i) investigations in simple models (Table 2),
(ii) investigations with operational-scale models, but with simulated observations (Table 3), and
(iii) as (ii), but with real data (Table 4).
The nomenclature used here is consistent with Figure 1, even though some articles use their own.

The main outcomes are that the 4D methods tend to be better than 3D methods, hybrid methods tend to be better than pure Var, ensemble, or EnVar methods, (and hybrids are more robust when N is small) but 4DEnVar still needs some development effort if it is to become better than En4DVar in all systems; only one study found hybrid 4DEnVar gave better performance than hybrid En4DVar (Gustafsson and Bojarova,

Table 2. Studies made of some of the techniques in this article applied to simplified models.

Article	Methods tested	Model	N	Best performance	Comments
Etherton and Bishop (2004)	3D-Var, hybrid En3DVar	2D turbulent barotropic vorticity	16, 64	Hybrid En3DVar	Model error included, and using sub-optimal ensemble
Hunt *et al.* (2004)	EnKF, 4DEnVar	Lorenz 96 (Lorenz, 1996)	40	4DEnVar	–
Wang *et al.* (2007a)	Hybrid OI, EnKF	Two-layer primitive equation	5, 20, 50	hybrid OI for small N	–
Liu *et al.* (2008)	3/4D-Var, EnKF, 4DEnVar	1D shallow-water	50	4DEnVar comparable to 4D-Var	4DEnVar cheaper than 4D-Var
Tian *et al.* (2008)	4D-Var, EnKF, 4DEnVar	NCAR Land (Oleson *et al.*, 2004), Lorenz 63 (Lorenz *et al.*, 1963)	60	4DEnVar	Based on 'Proper Orthogonal Decomposition'
Wang *et al.* (2009)	EnSRF, hybrid En3DVar	Two-layer primitive equation	50, 200	Hybrid En3DVar	Includes model error due to unresolved scales
Zhang *et al.* (2009)	4D-Var, EnKF, hybrid En4DVar	Lorenz 96	10, 40	Hybrid En4DVar	Hybrid less sensitive to window length and N
Fairbairn *et al.* (2014)	4D-Var, EnKF, En4DVar, 4DEnVar	Lorenz 2005 (Lorenz, 2005)	3–150	En4DVar especially for small N	4D-Var competitive when model is imperfect
Penny (2014)	LETKF, hybrid covariance, hybrid gain	Lorenz 96	2–40	Hybrid cov/ hybrid gain	Hybrids allow smaller N
Goodliff *et al.* (2015)	ETKF, ETKS, En4DVar, 4DEnVar, hybrid En4DVar	Lorenz 63	5, 10, 20, 50, 100	ETKS/ hybrid En4DVar	Hybrid En4DVar best for short window lengths
Poterjoy and Zhang (2015)	Hybrid En4DVar, hybrid 4DEnVar	Lorenz 96	5, 10, 20, 40	Hybrid En4DVar	Methods similar in data-dense areas

2014). However 4DEnVar is much cheaper to run than En4DVar so, if the performance of 4DEnVar could be improved through further work, it could then allow the cost savings in computation to be shifted to models with higher resolution and with more ensemble members. The results from the simplified systems are in general in agreement with those from the full systems, confirming the value of the simplified studies, and all studies have helped ultimately guide centres to decide on which methods to use operationally.

10. Summary of variational/ensemble methods used in NWP centres

It is often reported that most, if not all, NWP centres have the capability to combine their ensemble and variational systems. Table 5 lists the capabilities of a number of leading (inter)national NWP centres and research groups. Obviously this list is a snapshot and the true capabilities will change as developments progress.

11. Summary and outlook

Data assimilation needs to work well to serve NWP. State-of-the-art Var systems assume that background information obeys Gaussian statistics, with a covariance, $\mathbf{P}^{\mathrm{b}}$, estimated from a quasi-static representation, from an ensemble, or a combination of both. The use of ensemble information in Var systems using the raft of methods discussed in this article to give appropriate flow dependence of $\mathbf{P}^{\mathrm{b}}$, has been challenging, but largely successful.

This article is intended to be a pedagogical review of Var methods and the practical ways of introducing ensemble information into them to take advantage of the efficiency of Var and the flow-dependence of an ensemble, while minimizing their drawbacks (Buehner *et al.*, 2015b, and section 1). Broadly speaking, this is presently done in three ways.

1. Use the ensemble to recalibrate the (co)variances of the $\mathbf{B}_0$-matrix used as an approximation for $\mathbf{P}^{\mathrm{b}}$ in 4D-Var (section 5.1).
2. Use the ensemble to define $\mathbf{P}^{\mathrm{b}}$ by preconditioning the control vector on the ensemble itself (so-called pure EnVar methods). This describes the analysis increment as a linear combination of ensemble perturbations (where the control vector contains the coefficients) and implies a background-error covariance matrix equal to the sample covariance of the ensemble, $\mathbf{P}^{\mathrm{b}}_{(N)}$. This can be done using methods that exploit (En4DVar) or avoid (4DEnVar) the linear/adjoint forecast model (section 4).
3. Average $\mathbf{B}_0$ with $\mathbf{P}^{\mathrm{b}}_{(N)}$ (so-called hybrid EnVar). This is usually done with an augmented control vector comprising a part that is associated with the static part (as used in traditional Var) and a part that is associated with the ensemble (the same as that used in 1. here). Alternative methods are reviewed such as the a hybrid gain method, the EVIL method, and an ensemble reduced-rank Kalman filter (section 5.3).

These methods use the Var machinery to solve a DA problem and most require a separate ensemble system, which many NWP centres have anyway for ensemble prediction purposes. Practical methods that use ensemble information rely on localization to mitigate some of the effects of sampling error. The mathematical complexities that localization brings can mask the workings of the equations and so the methods are shown first without localization, but it is introduced later with two Schur product methods documented by Buehner (2005) (B05) and Lorenc (2003) (L03).

EnVar using 2. and 3. here has been shown to be fruitful and the key reasons are summarized as:

- (Hybrid) En4DVar and (hybrid) 4DEnVar are efficient ways of finding the deterministic (single) analysis.
- The methods provide flow-dependency to the $\mathbf{P}^{\mathrm{b}}$ used in the DA.

Table 3. Studies made of some of the techniques in this article applied to operational-scale models, but with simulated or severely reduced observations.

Article	Methods tested	Model	*N*	Best performance	Comments
Wang *et al.* (2008a)	3D-Var, hybrid En3DVar	WRF (Skamarock *et al.*, 2005)	50	Hybrid En3DVar	Benefit of hybrid especially over data-sparse regions
Liu *et al.* (2009)	3/4DEnVar	WRF	37	4DEnVar	Humidity particularly sensitive to sampling error
Buehner *et al.* (2010a)	3D-FGAT, 4D-Var, En3/4DVar, 4DEnVar	GEM (EC) (Cote *et al.*, 1998)	72, 96	–	Single obs./column expts.
Kleist and Ide (2015a)	3D-Var, hybrid 3DEnVar	NCEP GFS (NCEP, 2015)	80	Hybrid 3DEnVar	–
Kleist and Ide (2015b)	4D-Var, 4DEnVar, hybrid 4DEnVar	NCEP GFS	40, 80	Hybrid 4DEnVar	Improvement from 3D to 4D less than improvement in (2015a)

Table 4. Studies made of some of the techniques in this article applied to operational-scale models with real observations.

Article	Methods tested	Model	*N*	Best performance	Comments
Wang *et al.* (2008b)	3D-Var, hybrid En3DVar	WRF	50	Hybrid En3DVar	–
Buehner *et al.* (2010b)	3D-FGAT, 4D-Var, En3/4DVar, 4DEnVar	GEM (EC)	96	En4DVar	One month experiment
Hamill *et al.* (2011b)	3D-Var, EnKF, hybrid En3DVar	NCEP GFS	?	EnKF/ hybrid En3DVar	–
Wang (2011)	3D-Var, hybrid En3DVar	WRF	32	Hybrid En3DVar	Hybrid was able to adjust hurricane position
Clayton *et al.* (2012)	3/4D-Var, hybrid En3/4DVar	Global UM (Davies *et al.*, 2005)	23	Hybrid En4DVar	4D has marginal improvement over 3D in hybrid
Zhang and Zhang (2012)	4D-Var, EnKF, hybrid En4DVar	WRF	10, 40	Hybrid En4DVar	Period of summer convection
Buehner *et al.* (2013)	3/4D-Var, hybrid 3DEnVar, hybrid 4DEnVar	GEM (EC)	192	Hybrid 4DEnVar	Hybrid 4DEnVar is viable method
Kuhl *et al.* (2013)	4D-Var, hybrid En4DVar	NOGAPS (Hogan *et al.*, 1991)	80	Hybrid En4DVar	–
Liu and Xiao (2013)	3D-Var, 3D-FGAT, 3/4DEnVar	WRF	38	4DEnVar	Includes ensemble of En4DVar
Schwartz *et al.* (2013)	3D-Var, hybrid En3DVar	WRF	32	Hybrid En3DVar	Hybrid best for one outer loop, 3D-Var often best for three outer loops
Wang *et al.* (2013)	3D-Var, EnKF, En3DVar, hybrid En3DVar	NCEP GFS	80	En3DVar	Tested one-way and two-way coupling
Zhang *et al.* (2013)	3D-Var, EnKF, hybrid En3DVar, hybrid En4DVar	WRF	10, 20, 40, 80	Hybrid En4DVar	Hybrid En3DVar with half *N* comparable to EnKF
Pan *et al.* (2014)	3D-Var, hybrid En3DVar	WRF	40	Hybrid En3DVar	Tested one-way and two-way coupling
Poterjoy and Zhang (2014)	4D-Var, EnKF, hybrid En4DVar	WRF	60	Hybrid En4DVar	Tested period of tropical cyclogenesis and hurricane path
Gustafsson *et al.* (2014)	3D-Var, 4D-Var, hybrid En3DVar, hybrid En4DVar	HIRLAM Unden *et al.* (2002)	20, 40	Hybrid En4DVar	Marginal gain only
Gustafsson and Bojarova (2014)	4D-Var, hybrid En4DVar, hybrid 4DEnVar	HIRLAM	20	Hybrid 4DEnVar	–
Schwartz and Liu (2014)	3D-Var, EnSRF, hybrid En3DVar	WRF	50	Hybrid En3DVar	Hybrid best for precip. positions, but methods otherwise comparable
Schwartz *et al.* (2014)	3D-Var, EnSRF, hybrid En3DVar	WRF-Chem (Schwartz *et al.*, 2012)	50	Hybrid En3DVar	Aerosol assimilation
Wang and Lei (2014)	En3/4DVar	NCEP GFS	80	4DEnVar	–
Buehner *et al.* (2015a)	4D-Var, hybrid 4DEnVar	GEM (EC)	256	Hybrid 4DEnVar	Cost savings allow higher resolution
Caron *et al.* (2015)	4D-Var, hybrid 4DEnVar	Regional version of GEM (EC)	192, 256	Hybrid 4DEnVar	Hybrid 4DEnVar an order of magnitude cheaper than 4D-Var
Lorenc *et al.* (2015)	hybrid En3/4DVar, Hybrid 3/4DEnVar	Global UM	44	Hybrid En4DVar	–

- Although the En4DVar and 4DEnVar solutions are nominally equivalent to the ensemble mean of the analogous EnKF system (these methods and their hybrid counterparts allow an extension to the 4D domain), they allow access to the full range of observation operators, and allow other on-line features to be facilitated, namely variational bias correction and variational initialization.
- In (hybrid) En4DVar, the linearized forecast model (and its adjoint) are used (as in 4D-Var), but this is avoided in (hybrid) 4DEnVar. This is the purpose of developing the latter as the linearization of the forecast model is time-consuming to construct and maintain, is costly to run, and is limited by the nature of the physical processes, especially parametrizations with switches.

Table 5. Current deterministic DA capabilities at a selection of NWP centres and groups worldwide according to published material.

System/ Group	Model	Current capability	Local-ization	N	Key references and comments
DWD	ICON (global) (Zangl *et al.*, 2015)	Hybrid 3DEnVar	B05	40	Rhodin (2015), specifications based on unpublished sources.
ECMWF	IFS (global) (ECMWF, 2015)	Recalibration of vars and corrs in 4D-Var	–	25	Bonavita *et al.* (2016), see section 5.1
EC	GEM (global) (Cote *et al.*, 1998)	Hybrid 4DEnVar	B05	256	Buehner *et al.* (2015a), Assim./forecast system known as GDPS
EC	GEM (regional)	Hybrid 4DEnVar	B05	256	Caron *et al.* (2015), Assim./forecast system known as RDPS
HIRLAM	HIRLAM (regional) (Unden *et al.*, 2002)	hybrid En4DVar, hybrid 4DEnVar	L03	40	Gustafsson *et al.* (2014); Gustafsson and Bojarova (2014)
Météo-France	Arpège (global) (Courtier *et al.*, 1991)	Recalibration of vars in 4D-Var	–	25	Berre *et al.* (2007); Raynaud *et al.* (2011, 2012), Berre *et al.* (2015)
Met Office	UM (global) (Davies *et al.*, 2005)	Hybrid En4DVar	L03	23	Clayton *et al.* (2012)
NRL	NOGAPS (global) (Hogan *et al.*, 1991)	Hybrid En4DVar	B05	80	Kuhl *et al.* (2013); Bishop *et al.* (2011)
NCEP	GFS	Hybrid 4DEnVar	L03	80	Wang and Lei (2014); Kleist and Ide (2015a,2015b)
NCEP	WRF (Skamarock *et al.*, 2005), WRF-Chem (Schwartz *et al.*, 2012)	4DEnVar, hybrid 3D-Var	B05, L03	50, 37	Wang *et al.* (2008a); Liu *et al.* (2009); Wang (2010) and Schwartz *et al.* (2014)
PSU	WRF	Hybrid En4DVar	L03	40	Zhang and Zhang (2012)
UoSF	WRF	4DEnVar	B05	38	Liu *et al.* (2009)

Arpège (Action de Recherche Petite Echelle Grande Echelle), DWD (Deutscher Wetterdienst), EC (Environment Canada), ECMWF (European Centre for Medium Range Weather Forecasting), GEM (Global Environmental Multi-scale), GDPS (Global Deterministic Prediction System), GFS (Global Forecasting System), HIRLAM (High Resolution Local Area Modelling), ICON (ICOsahedral Non-hydrostatic model), IFS (Integrated Forecasting System), NCAR (National Center for Atmospheric Research), NCEP (National Centers for Environmental Prediction), NOGAPS (Navy Operational Global Atmospheric Prediction System), NRL (Navy Research Laboratory), PSU (Pennsylvania State Univ.), UM (Unified Model), UoSF (Univ. of South Florida), WRF (Weather Research and Forecasting model).

- As these methods use information from an ensemble, there is still sensitivity to the ensemble size, N, especially when it is small, but hybrid methods have been shown to reduce the sensitivity to N.
- (Hybrid) EnVar methods are competitive with, or outperform, the pure Var or ensemble counterparts.

There are outstanding issues that still need to be solved. For instance in hybrid 4DEnVar the 4D ensemble is averaged with 3D-FGAT and not with 4D-Var because the latter needs the linearized model (section 7.4). Possible ways forward involve extensions to $\mathbf{U}$ (the control variable transform) or to reintroduce a linear model, albeit highly simplified (section 7.4.1). Furthermore in (hybrid) 4DEnVar the way that 4D localization is done is thought to be crucial to its effectiveness. It is relatively straightforward to do 4D localization with functions that are separable in space and time, but there are reasons why this is not appropriate (features that are together at $t = 0$ – so unaffected by localization – would remain correlated at $t > 0$, yet may be further apart, and so would be wrongly damped by localization). Such a problem may be dealt with by advecting the localization with similar simplified linear model or by using an adaptive scheme (e.g. Bishop and Hodyss, 2007,2009a,2009b,2011). How to account for model error in (hybrid) 4DEnVar is also an ongoing problem (section 8), especially if it introduces the need for a model error covariance matrix (which brings us back to a covariance calibration problem, as for $\mathbf{B}_0$).

There is scope to test the methods covered in this review to situations where pure Var can struggle. One application is quantitative precipitation forecasting at the convective scale. There are already some encouraging results in this area: e.g. hybrid En4DVar can outperform pure ensemble and Var methods in summer convection (Zhang and Zhang, 2012), and hybrid En3DVar is better than pure ensemble and Var methods for analysing precipitation positions (Schwartz and Liu, 2014). Furthermore hybrid 3DEnVar (unlike 3D-Var) can improve the position of a hurricane (Wang, 2011). However, there are some less encouraging studies showing that, for example, humidity is particularly sensitive to sampling error and is not improved with 4DEnVar over 3DEnVar (Liu *et al.*, 2009). Operational centres like the Met Office use EnVar systems for their global models, but have yet to develop similar systems for their regional models. It remains an important question how much these techniques can improve analyses and forecasts in their high-resolution regional models. Environment Canada though has already shown that 4DEnVar can lead to better forecasts of moisture-related quantities like precipitable water and equitable threat score for precipitation (Caron *et al.*, 2015). It should also be noted that all of the methods may be suitable for application to other geophysical problems like coupled systems and estimating surface fluxes of chemical species.

Finally a note on nomenclature. As the number of methods grows, it becomes important to label each succinctly and accurately. Lorenc (2013) has attempted to do this, and authors have tended to follow. We support the view that the 'hybrid' label should be used only for methods that combine ensemble and static aspects in their background errors, but we do think that it should be necessary to specify this when a method does use information from both sources (unlike point 5. in Lorenc, 2013)); see Figure 1. It is also useful to distinguish between 3D and 4D pure ensemble methods. In this case it is sensible to specify 4D ensemble methods as smoothers (EnKS) rather than as filters (EnKF). To add to the complexity, there are also methods that run ensembles of 4D-Var (Belo Pereira and Berre, 2006; Berre *et al.*, 2007, 2009; Isaksen *et al.*, 2010) or of 4DEnVar (Fairbairn *et al.*, 2014) and a nomenclature is needed here. One suggestion is to use En-4DVar and En-4DEnVar respectively to indicate that the 'En' (separated from the main method with a hyphen) represents independent runs of the system specified.

Acknowledgements

The author is supported by NERC via the NCEO. I am constantly learning new things about DA from my colleagues, but I would especially like to thank people who, over the years, have given inspiring talks or I have had interesting conversations with concerning the subject matter of this article, namely Javier Amezcua, Jeff Anderson, Dale Barker, Craig Bishop, Neil Bowler, Mark Buehner, Jean-Francois Caron, Mike Cullen, Martin Ehrendorfer, David Fairbairn, Jonathan Flowerdew, Mike Fisher, Michael Goodliff, Gordon Inverarity, Daryl Kleist, Amos Lawless, David Livings, Andrew Lorenc, Stefano Migliorini, Chris Snyder, Peter Jan van Leeuwen, Marek Wlasak and Fuqing Zhang. Thanks also to the anonymous reviewers of this article for providing important comments.

Appendices

Appendix A: The vector derivative

Consider a quadratic function of the n-element vector $\mathbf{x}$:

$$J(\mathbf{x}) = \left\|\mathbf{H}\mathbf{x}+\mathbf{y}\right\|^2_{\mathbf{S}^{-1}} = \frac{1}{2}(\mathbf{H}\mathbf{x}+\mathbf{y})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{H}\mathbf{x}+\mathbf{y}), \tag{A1}$$

where $\mathbf{y}$ (p-elements), $\mathbf{H}$ ($p\times n$-elements) and the symmetric matrix $\mathbf{S}$ ($p\times p$-elements). Derivatives of such a function appear frequently, and so we derive a general result. The vector derivative is the following n-element vector:

$$\nabla_{\mathbf{x}}J = \begin{pmatrix} \partial J/\partial \mathbf{x}_1 \\ \vdots \\ \partial J/\partial \mathbf{x}_n \end{pmatrix}, \tag{A2}$$

where $\mathbf{x}_i$ is the ith component of $\mathbf{x}$. First decompose (A1) into a function of its components $\mathbf{x}_i$:

$$J(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\mathbf{x}_l+\mathbf{y}_i\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\mathbf{x}_k+\mathbf{y}_j\right). \tag{A3}$$

Differentiating this mechanically with respect to one arbitrary element $\mathbf{v}_m$ gives, by the product rule,

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{x}_m} =& \frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\frac{\partial \mathbf{x}_l}{\partial \mathbf{x}_m}\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\mathbf{x}_k+\mathbf{y}_j\right) \\
&+\frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\mathbf{x}_l+\mathbf{y}_i\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_p}\right) \\
=& \frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\delta_{lm}\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\mathbf{x}_k+\mathbf{y}_j\right) \\
&+\frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\mathbf{x}_l+\mathbf{y}_i\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\delta_{kp}\right) \\
=& \frac{1}{2}\sum_{i=1}^{p}\mathbf{H}_{im}\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\mathbf{x}_k+\mathbf{y}_j\right) \\
&+\frac{1}{2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\mathbf{H}_{il}\mathbf{x}_l+\mathbf{y}_i\right)\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\mathbf{H}_{jp}.
\end{aligned} \tag{A4}$$

In the second line of the last expression, re-index the summations as $i \longleftrightarrow j$ and $l \to k$. Noting that $\left(\mathbf{S}^{-1}\right)_{ij} = \left(\mathbf{S}^{-1}\right)_{ji}$, this shows that the two lines in the last expression are identical, so:

$$\frac{\partial J}{\partial \mathbf{x}_m} = \sum_{i=1}^{p}\mathbf{H}_{im}\sum_{j=1}^{p}\left(\mathbf{S}^{-1}\right)_{ij}\left(\sum_{k=1}^{n}\mathbf{H}_{jk}\mathbf{x}_k+\mathbf{y}_j\right). \tag{A5}$$

This can be confirmed to be the decomposition of the mth element of the matrix expression:

$$\nabla_{\mathbf{x}}J = \mathbf{H}^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{H}\mathbf{x}+\mathbf{y}). \tag{A6}$$

Appendix B: The matrix derivative

Consider a quadratic function of the $n\times N$ matrix $\mathbf{V}$:

$$J(\mathbf{V}) = \frac{1}{2}\left\{\mathbf{H}\left(\mathbf{X}\circ[\mathbf{U}\mathbf{V}]\right)\mathbf{1}_N+\mathbf{y}\right\}^{\mathrm{T}}\mathbf{S}^{-1}\left\{\mathbf{H}\left(\mathbf{X}\circ[\mathbf{U}\mathbf{V}]\right)\mathbf{1}_N+\mathbf{y}\right\}, \tag{B1}$$

where $\mathbf{y}$ (p-elements), $\mathbf{H}$ ($p\times n$-elements), $\mathbf{X}$ ($n\times N$-elements), $\mathbf{U}$ ($n\times n$-elements), the symmetric matrix $\mathbf{S}$ ($p\times p$-elements), and $\mathbf{1}_N$ (N-elements, each of value 1). This is the structure of cost functions that use the L03 formulation of localizing ensemble-derived background-error covariances ((66) in section 6.3). The matrix derivative is defined as

$$\nabla_{\mathbf{V}}J = \begin{pmatrix} \partial J/\partial \mathbf{V}_{11} & \cdots & \partial J/\partial \mathbf{V}_{1N} \\ \vdots & \ddots & \vdots \\ \partial J/\partial \mathbf{V}_{n1} & \cdots & \partial J/\partial \mathbf{V}_{nN} \end{pmatrix}, \tag{B2}$$

where $\mathbf{V}_{ij}$ is the ijth element of $\mathbf{V}$. In order to derive a matrix expression for this gradient, we use the chain rule. Let $\mathbf{v} = \{\mathbf{X}\circ[\mathbf{U}\mathbf{V}]\}\mathbf{1}_N$, where $\mathbf{v}$ has n elements. $\nabla_{\mathbf{V}}J$ can be found using the vector derivative result in Appendix A:

$$\nabla_{\mathbf{v}}J = \mathbf{H}^{\mathrm{T}}\mathbf{S}^{-1}\left(\mathbf{H}\mathbf{v}+\mathbf{y}\right). \tag{B3}$$

The chain rule allows the derivative with respect to $\mathbf{Z}$ to be found:

$$\frac{\partial J}{\partial \mathbf{V}_{ij}} = \sum_{l=1}^{n}\frac{\partial \mathbf{v}_l}{\partial \mathbf{V}_{ij}}\frac{\partial J}{\partial \mathbf{v}_l}. \tag{B4}$$

$\partial J/\partial \mathbf{v}_l$ which appears in (B4) is the lth component of (B3), and $\partial \mathbf{v}_l/\partial \mathbf{V}_{ij}$ can be found by first expanding out the definition of $\mathbf{v}$ in terms of $\mathbf{V}$ given above:

$$\mathbf{v}_l = \sum_{j'=1}^{N}\mathbf{X}_{lj'}\sum_{i'=1}^{n}\mathbf{U}_{li'}\mathbf{V}_{i'j'}, \tag{B5}$$

$$\begin{aligned}
\text{so}\quad \frac{\partial \mathbf{v}_l}{\partial \mathbf{V}_{ij}} &= \sum_{j'=1}^{N}\mathbf{X}_{lj'}\sum_{i'=1}^{n}\mathbf{U}_{li'}\frac{\partial \mathbf{V}_{i'j'}}{\partial \mathbf{V}_{ij}} \\
&= \sum_{j'=1}^{N}\mathbf{X}_{lj'}\sum_{i'=1}^{n}\mathbf{U}_{li'}\delta_{ii'}\delta_{jj'} = \mathbf{X}_{lj}\mathbf{U}_{li}.
\end{aligned} \tag{B6}$$

Substituting this into the chain rule gives:

$$\frac{\partial J}{\partial \mathbf{V}_{ij}} = \sum_{l=1}^{n}\mathbf{X}_{lj}\mathbf{U}_{li}\frac{\partial J}{\partial \mathbf{v}_l}, \tag{B7}$$

which can be confirmed to be the ijth element of the matrix expression:

$$\nabla_{\mathbf{V}}J = \mathbf{U}^{\mathrm{T}}\left[\mathbf{X}\circ\left(\nabla_{\mathbf{v}}J \quad \cdots \quad \nabla_{\mathbf{v}}J\right)\right], \tag{B8}$$

where $\left(\nabla_{\mathbf{v}}J \quad \cdots \quad \nabla_{\mathbf{v}}J\right)$ is the $n\times N$ matrix of repeated columns of $\nabla_{\mathbf{v}}J$.

Appendix C: The variational form of the Ensemble Kalman Smoother

The correspondence between the EnKS formula (29) and the minimizing state of (35) is demonstrated here. The gradient of (35) with respect to $\boldsymbol{\chi}_{\text{ens}}$ is given in (37). Setting this is [AUTHOR: as?] zero for the minimum gives the following solution:

$$\boldsymbol{\chi}_{\text{ens}} = \left[\mathbf{I} + \left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right]^{-1}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{d}}. \tag{C1}$$

(Recall the definitions $\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}$ and $\underline{\mathbf{H}}_{\mathbf{M}}$ in (31), $\underline{\mathbf{d}}$ is defined after (35), $\underline{\mathbf{R}}$ is the time-ordered observation-error covariance matrix, and $\mathbf{X}^{\text{b}}$ is the matrix whose columns comprise the normalized ensemble of 3D perturbations, $(\mathbf{x}^{\text{b}}_{(i)} - \overline{\mathbf{x}^{\text{b}}})/\sqrt{N-1}$). The model space increment is found from (36):

$$\delta\mathbf{x} = \mathbf{X}^{\text{b}}\left[\mathbf{I} + \left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right]^{-1}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{d}}. \tag{C2}$$

The standard Sherman–Morison–Woodbury identity (e.g. Appendix A of Ehrendorfer, 2007):

$$(\mathbf{B}^{-1} + \mathbf{H}^{\text{T}}\mathbf{R}^{-1}\mathbf{H})\mathbf{B}\mathbf{H}^{\text{T}} = \mathbf{H}^{\text{T}}\mathbf{R}^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\text{T}})$$

applied to (C2) with $\mathbf{B} \rightarrow \mathbf{I}$, $\mathbf{H} \rightarrow \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}$, and $\mathbf{R} \rightarrow \underline{\mathbf{R}}$ gives:

$$\delta\mathbf{x} = \mathbf{X}^{\text{b}}\mathbf{X}^{\text{b}\text{T}}\underline{\mathbf{H}}_{\mathbf{M}}^{\text{T}}\left(\underline{\mathbf{R}} + \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\mathbf{X}^{\text{b}\text{T}}\underline{\mathbf{H}}_{\mathbf{M}}^{\text{T}}\right)^{-1}\underline{\mathbf{d}}. \tag{C3}$$

The EnKS (29) produces an ensemble of analyses. Taking the mean gives:

$$\overline{\mathbf{x}^{\text{a}}} - \overline{\mathbf{x}^{\text{b}}} = \mathbf{X}^{\text{b}}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\left\{\underline{\mathbf{R}} + \underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\left(\underline{\mathbf{H}}_{\mathbf{M}}\mathbf{X}^{\text{b}}\right)^{\text{T}}\right\}^{-1} \times \left\{\underline{\mathbf{y}}^{\text{o}} - \overline{\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\text{b}})}\right\}. \tag{C4}$$

This shows that, as long as the ensemble mean of the model observations, $\overline{\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\text{b}})}$, is close to the model observations of the deterministic background, $\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}^{\text{b}})$, then the minimization problem (35) is equivalent to the mean state of the EnKS (29). We are taking $\mathbf{x}^{\text{b}} = \overline{\mathbf{x}^{\text{b}}}$ (the ensemble mean) so the above result is true in the special case when $\underline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}$ is linear.

Appendix D: The 'B05' method of localizing covariances – the matrix of 'localized' ensemble members

A sample of N background perturbations (each of n-elements and divided by $\sqrt{N-1}$) are held in the matrix $\mathbf{X}^{\text{b}}$ and a sample of M 'localization' perturbations – see below. Each of n-elements divided by $\sqrt{M-1}$ are held in the matrix $\mathbf{U}^{\mathbf{C}}$. [AUTHOR: Is this the intended meaning?] The covariances of $\mathbf{X}^{\text{b}}$ and $\mathbf{U}^{\mathbf{C}}$ are defined as:

$$\mathbf{P}^{\text{b}}_{(N)} = \mathbf{X}^{\text{b}}\mathbf{X}^{\text{b}\text{T}}, \tag{D1}$$

$$\mathbf{C} = \mathbf{U}^{\mathbf{C}}\mathbf{U}^{\mathbf{C}\text{T}}. \tag{D2}$$

What is the matrix of effective ensemble perturbations, $\widehat{\mathbf{X}}^{\text{b}}$, whose covariance is the Schur product $\mathbf{C} \circ \mathbf{P}^{\text{b}}_{(N)}$? We will show that this is the following $n \times NM$ matrix:

$$\widehat{\mathbf{X}}^{\text{b}} = \frac{1}{\sqrt{N-1}}\left(\text{diag}(\delta\mathbf{x}^{\text{b}}_{(1)})\mathbf{U}^{\mathbf{C}}, \ldots, \text{diag}(\delta\mathbf{x}^{\text{b}}_{(N)})\mathbf{U}^{\mathbf{C}}\right), \tag{D3}$$

where $\delta\mathbf{x}^{\text{b}}_{(i)}/\sqrt{N-1}$ is the ith column of $\mathbf{X}^{\text{b}}$ and the diag operator gives an $n \times n$ diagonal matrix whose diagonal elements comprise the vector argument. There are NM columns of $\widehat{\mathbf{X}}^{\text{b}}$, which may be thought of as NM effective ensemble members whose covariance has the desired property. Evaluating $\widehat{\mathbf{X}}^{\text{b}}\widehat{\mathbf{X}}^{\text{b}\text{T}}$ from (D3) gives:

$$\begin{aligned}
\widehat{\mathbf{X}}^{\text{b}}\widehat{\mathbf{X}}^{\text{b}\text{T}} &= \frac{1}{N-1}\left(\text{diag}(\delta\mathbf{x}^{\text{b}}_{(1)})\mathbf{U}^{\mathbf{C}}, \ldots, \text{diag}(\delta\mathbf{x}^{\text{b}}_{(N)})\mathbf{U}^{\mathbf{C}}\right) \\
&\quad \times \begin{pmatrix} \mathbf{U}^{\mathbf{C}\text{T}}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(1)}) \\ \vdots \\ \mathbf{U}^{\mathbf{C}\text{T}}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(N)}) \end{pmatrix} \\
&= \frac{1}{N-1}\sum_{k=1}^{N}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})\mathbf{U}^{\mathbf{C}}\mathbf{U}^{\mathbf{C}\text{T}}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)}) \\
&= \frac{1}{N-1}\sum_{k=1}^{N}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})\mathbf{C}\,\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)}).
\end{aligned} \tag{D4}$$

The i,jth element of (D4) is given below, which is shown to be the i,jth element of $\mathbf{P}^{\text{b}}_{(N)} \circ \mathbf{C}$:

$$\begin{aligned}
&\left(\widehat{\mathbf{X}}^{\text{b}}\widehat{\mathbf{X}}^{\text{b}\text{T}}\right)_{ij} \\
&= \frac{1}{N-1}\sum_{k=1}^{N}\sum_{p=1}^{n}\sum_{q=1}^{n}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})_{ip}\mathbf{C}_{pq}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})_{qj} \\
&= \frac{1}{N-1}\sum_{k=1}^{N}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})_{ii}\mathbf{C}_{ij}\text{diag}(\delta\mathbf{x}^{\text{b}}_{(k)})_{jj} \\
&= \frac{1}{N-1}\left[\sum_{k=1}^{N}(\delta\mathbf{x}^{\text{b}}_{(k)})_i(\delta\mathbf{x}^{\text{b}}_{(k)})_j\right]\mathbf{C}_{ij} \\
&= \mathbf{P}^{\text{b}}_{ij}\mathbf{C}_{ij}. \\
&\therefore \widehat{\mathbf{X}}^{\text{b}}\widehat{\mathbf{X}}^{\text{b}\text{T}} = \mathbf{P}^{\text{b}} \circ \mathbf{C}.
\end{aligned} \tag{D5}$$

Appendix E: The 'L03' method of localizing covariances

This appendix shows that the CVT given in (68):

$$\delta\mathbf{x} = \left\{\mathbf{X}^{\text{b}} \circ (\mathbf{U}^{\mathbf{C}}\widehat{\mathbf{V}}\right\}\mathbf{1}_N, \tag{E1}$$

implies a background-error covariance matrix that is localized. For N ensemble members and a state size n, here $\mathbf{X}^{\text{b}}$ ($n \times N$-elements) is the matrix of scaled background deviations (i.e. $\mathbf{X}^{\text{b}}_{ji} = (\mathbf{x}^{\text{b}}_{(i)} - \overline{\mathbf{x}^{\text{b}}})_j/\sqrt{N-1}$), $\mathbf{U}^{\mathbf{C}}$ ($n \times n$-elements) is a square-root of localization matrix $\mathbf{C}$, $\widehat{\mathbf{V}}$ ($n \times N$-elements) is the matrix whose columns are the mutually uncorrelated control vectors $\widehat{\boldsymbol{\chi}}_{(j)}$ (i.e. $\widehat{\mathbf{V}}_{ij} = (\widehat{\boldsymbol{\chi}}_{(j)})_i$ –the ith component of the jth control vector), and $\mathbf{1}_N$ is the N-element column vector of 1s. The kth element of $\delta\mathbf{x}$ from (E1) is:

$$\begin{aligned}
\delta\mathbf{x}_k &= \sum_{m=1}^{N}\mathbf{X}^{\text{b}}_{km}\sum_{l=1}^{n}\mathbf{U}^{\mathbf{C}}_{kl}\widehat{\mathbf{V}}_{lm} \\
&= \frac{1}{\sqrt{N-1}}\sum_{m=1}^{N}(\delta\mathbf{x}^{\text{b}}_{(m)})_k\sum_{l=1}^{n}\mathbf{U}^{\mathbf{C}}_{kl}(\widehat{\boldsymbol{\chi}}_{(m)})_l.
\end{aligned}$$

Calculating the covariance between elements k and k' of $\delta\mathbf{x}$:

$$\begin{aligned}
&\langle\delta\mathbf{x}_k\delta\mathbf{x}_{k'}\rangle \\
&= \frac{1}{N-1}\left\langle\left(\sum_{m=1}^{N}(\delta\mathbf{x}^{\text{b}}_{(m)})_k\sum_{l=1}^{n}\mathbf{U}^{\mathbf{C}}_{kl}(\widehat{\boldsymbol{\chi}}_{(m)})_l\right)\right. \\
&\qquad \left.\left(\sum_{m'=1}^{N}(\delta\mathbf{x}^{\text{b}}_{(m')})_{k'}\sum_{l'=1}^{n}\mathbf{U}^{\mathbf{C}}_{k'l'}(\widehat{\boldsymbol{\chi}}_{(m')})_{l'}\right)\right\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N-1}\sum_{m=1}^{N}\sum_{m'=1}^{N}(\delta\mathbf{x}^{\mathrm{b}}_{(m)})_k(\delta\mathbf{x}^{\mathrm{b}}_{(m')})_{k'} \\
&\qquad \times \sum_{l=1}^{n}\sum_{l'=1}^{n}\mathbf{U}^{\mathrm{C}}_{kl}\mathbf{U}^{\mathrm{C}}_{k'l'}\langle(\boldsymbol{\chi}_{(m)})_l(\boldsymbol{\chi}_{(m')})_{l'}\rangle \\
&= \frac{1}{N-1}\sum_{m=1}^{N}\sum_{m'=1}^{N}(\delta\mathbf{x}^{\mathrm{b}}_{(m)})_k(\delta\mathbf{x}^{\mathrm{b}}_{(m')})_{k'} \\
&\qquad \times \sum_{l=1}^{n}\sum_{l'=1}^{n}\mathbf{U}^{\mathrm{C}}_{kl}\mathbf{U}^{\mathrm{C}}_{k'l'}\delta_{mm'}\delta_{ll'} \\
&= \frac{1}{N-1}\sum_{m=1}^{N}(\delta\mathbf{x}^{\mathrm{b}}_{(m)})_k(\delta\mathbf{x}^{\mathrm{b}}_{(m)})_{k'}\sum_{l=1}^{n}\mathbf{U}^{\mathrm{C}}_{kl}\mathbf{U}^{\mathrm{C}\,\mathrm{T}}_{lk'} \\
&= \mathbf{P}^{\mathrm{b}}_{(N)\,kk'}\mathbf{C}_{kk'}.
\end{aligned}
$$

The left-hand side is matrix element k,k' of $\langle\delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}}\rangle$ and the right-hand side of the last line is matrix element k,k' of $\mathbf{P}^{\mathrm{b}}_{(N)}\circ\mathbf{C}$. This shows that the error covariance matrix implied by the CVT is the error covariance matrix sampled from the forecast ensemble, $\mathbf{P}^{\mathrm{b}}_{(N)}$ localized with $\mathbf{C}$.

References

Amezcua J, Goodliff M, van Leeuwen PJ. 2017. A weak-constraint 4DEnsembleVar. Part I: formulation and simple model experiments. *Tellus* **69A**: 1271564, doi: 10.1080/16000870.2016.1271564.

Anderson JL. 2001. An ensemble adjustment filter for data assimilation. *Mon. Weather Rev.* **129**: 2884–2903.

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**: 2741–2758.

Auligné T, Ménétrier B, Lorenc AC, Buehner M. 2016. Ensemble-variational integrated localised data assimilation. *Mon. Weather Rev.* **144**: 3677–3696, doi: 10.1175/MWR-D-15-0252.1.

Baker LH, Rudd AC, Migliorini S, Bannister RN. 2014. Representation of model error in a convective-scale ensemble prediction system. *Nonlin. Proc. Geophys.* **21**: 19–39.

Bannister RN. 2007. Can wavelets improve the representation of forecast error covariances in variational data assimilation? *Mon. Weather Rev.* **135**: 387–408.

Bannister RN. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast-error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**: 1971–1996.

Bannister RN. 2015. How is the balance of a forecast ensemble affected by adaptive and non-adaptive localization schemes? *Mon. Weather Rev.* **143**: 3680–3699.

Barker DM, Huang W, Guo Y-R, Bourgeois AJ, Xiao QN. 2004. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Weather Rev.* **132**: 897–914.

Beck A, Ehrendorfer M. 2005. Singular-vector-based covariance propagation in a quasigeostrophic assimilation system. *Mon. Weather Rev.* **133**: 1295–1310.

Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background-error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.

Benedetti A, Stephens GL, Tomislava V. 2003. Variational assimilation of radar reflectivities in a cirrus model. I: Model description and adjoint sensitivity studies. *Q. J. R. Meteorol. Soc.* **129**: 277–300.

Berner J, Ha S-Y, Hacker JP, Fournier A, Snyder C. 2011. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Weather Rev.* **139**: 1972–1995.

Berre L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Weather Rev.* **128**: 644–667.

Berre L, Stefănescu SE, Belo Pereira M. 2006. The representation of the analysis effect in three error simulation techniques. *Tellus* **58A**: 196–209.

Berre L, Pannekoucke O, Desroziers G, Stefănescu SE, Chapnik B, Raynaud L. 2007 'A variational assimilation ensemble and the spatial filtering of its error covariances: Increase of sample size by local spatial averaging'. In *Proceedings of the ECMWF Workshop on Flow-Dependent Aspects of Data Assimilation*, 11–13 June 2007. ECMWF: Reading, UK.

Berre L, Desroziers G, Raynaud L, Montroty R, Gibier F. 2009 'Use of consistent operational ensemble variational assimilation to estimate and diagnose analysis and background error covariances'. In *Proceedings of the ECMWF Workshop on Diagnostics of Data Assimilation System Performance*, 15–17 June 2009. ECMWF: Reading, UK.

Berre L, Varella H, Desroziers G. 2015. Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Q. J. R. Meteorol. Soc.* **141**: 2803–2812, doi: 10.1002/qj.2565.

Biegler LT. 1997. *Large-Scale Optimization with Applications: Part I: Optimization in Inverse Problems and Design*. Springer-Verlag: New York, NY.

Bishop CH, Hodyss D. 2007. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 2029–2044.

Bishop CH, Hodyss D. 2009a. Ensemble covariances adaptively localized with ECO-RAP, Part 1: Tests on simple error models. *Tellus* **61A**: 84–96.

Bishop CH, Hodyss D. 2009b. Ensemble covariances adaptively localized with ECO-RAP, Part 2: A strategy for the atmosphere. *Tellus* **61A**: 97–111.

Bishop CH, Hodyss D. 2011. Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Weather Rev.* **139**: 1241–1255.

Bishop CH, Satterfield EA. 2013. Hidden error variance theory. Part I: Exposition and analytic model. *Mon. Weather Rev.* **141**: 1454–1468.

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather Rev.* **129**: 420–436.

Bishop CH, Hodyss D, Steinle P, Sims H, Clayton AM, Lorenc AC, Barker DM, Buehner M. 2011. Efficient ensemble covariance localization in variational data assimilation. *Mon. Weather Rev.* **139**: 573–580.

Bonavita M, Torrisi L, Marcucci F. 2008. The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Q. J. R. Meteorol. Soc.* **134**: 1733–1744.

Bonavita M, Raynaud L, Isaksen L. 2011. Estimating background-error variances with the ECMWF ensemble of data assimilations system: Some effects of ensemble size and day-to-day variability. *Q. J. R. Meteorol. Soc.* **137**: 423–434, doi: 10.1002/qj.756.

Bonavita M, Isaksen L, Hólm EV. 2012. On the use of EDA background-error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.* **138**: 1540–1559, doi: 10.1002/qj.1899.

Bonavita M, Hólm EV, Isaksen L, Fisher M. 2016. The evolution of the ECMWF hybrid data assimilation system. *Q. J. R. Meteorol.Soc.* **142**: 287–303, doi: 10.1002/qj.2652.

Bouttier F, Vié B, Nuissier O, Raynaud L. 2012. Impact of stochastic physics in a convection-permitting ensemble. *Mon. Weather Rev.* **140**: 3706–3721.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 703–722.

Buehner M. 2005. Ensemble derived stationary and flow dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**: 1013–1043.

Buehner M. 2010. Evaluation of spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Weather Rev.* **140**: 617–636.

Buehner M, Charron M. 2007. Spectral and spatial localization of background-error correlations for data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 615–630.

Buehner M, Shlyaeva A. 2015. Scale-dependent background-error covariance localisation. *Tellus* **67A**: 28027, doi: 10.3402/tellusa.v67.28027.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Weather Rev.* **138**: 1550–1566.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Weather Rev.* **138**: 1567–1586, doi: 10.1175/2009MWR3158.1.

Buehner M, Morneau J, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlin. Proc. Geophys.* **20**: 669–682.

Buehner M, McTaggart-Cowan R, Beaulne A, Charette C, Garand L, Heilliette S, Lapalme E, Laroche S, Macpherson SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. *Mon. Weather Rev.* **143**: 2532–2559, doi: 10.1175/MWR-D-14-00354.1.

Buehner M, Gelaro R, Alves O, Dee DP, Kalnay E, van Leeuwen PJ, Lorenc AC, Miyoshi T, Saunders R, Schiller A. 2015b. Data assimilation methods and applications. Chapter 3 in *Seamless Prediction of the Earth System: From Minutes to Months*, Jones fnmS, Ruti PM. (eds.): 37–60. World Meteorological Organization: Geneva, Switzerland.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**: 2887–2908.

Burgers G, van Leeuwen P-J, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**: 1719–1724.

Caron J-F, Milewski T, Buehner M, Fillion L, Reszka M, Macpherson S, St-James J. 2015. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part II: The regional system. *Mon. Weather Rev.* **143**: 2560–2580, doi: 10.1175/MWR-D-14-00353.1.

Clayton AM, Lorenc AC, Barker DM. 2012. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**: 1445–1461, doi: 10.1002/qj.2054.

Côté J, Gravel S, Méthot A, Patioine A, Roch M, Staniforth A. 1998. The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Weather Rev.* **126**: 1373–1395.

Courtier P, Freydier C, Geleyn J-F, Rabier F, Rochas M. 1991. 'The Arpège project at Météo-France'. In *Proceedings of the ECMWF Workshop on Numerical Methods in Atmospheric Models*, 9–13 September 1991. ECMWF: Reading, UK.

Courtier P, Thepaut J-N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1387.

Courtier P, Andersson E, Heckley W, Pailleux J, Vasiljevic D, Hollingsworth A, Fisher M, Rabier F. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part I: Formulation. *Q. J. R. Meteorol. Soc.* **124**: 1783–1807.

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press: Cambridge, UK.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Deckmyn A, Berre L. 2005. A wavelet approach to representing background-error covariances in a limited-area model. *Mon. Weather Rev.* **133**: 1279–1294.

Dee DP. 2002. 'An adaptive scheme for cycling background-error variances during data assimilation'. In *Proceedings of ECMWF/GEWEX Workshop on Humidity Analysis*, 8–11 July 2002. ECMWF: Reading, UK.

Derber J, Bouttier F. 1999. A reformulation of the background-error covariance in the ECMWF global data assimilation system. *Tellus* **51A**: 195–221.

Desroziers G, Berre L. 2012. Accelerating and parallelizing minimizations in ensemble and deterministic variational assimilations. *Q. J. R. Meteorol. Soc.* **138**: 1599–1610.

Desroziers G, Camino J-T, Berre L. 2014. 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. *Q. J. R. Meteorol. Soc.* **140**: 2097–2110, doi: 10.1002/qj.2325.

ECMWF. 2015. 'Documentation for the integrated forecasting system'. https://software.ecmwf.int/wiki/display/IFS/\unhbox\voidb@x\hboxOfficial+IFS+Documentation (accessed 12 January 2017).

Ehrendorfer M. 2007. A review of issues in ensemble-based Kalman filtering. *Meteorol. Z.* **16**: 795–818.

El Akkraoui A, Trémolet Y, Todling R. 2013. Preconditioning of variational data assimilation and the use of a bi-conjugate gradient method. *Q. J. R. Meteorol.Soc.* **130**: 731–741.

Emerick A. 2012. *History Matching and Uncertainty Characterization*. Lambert Academic Publishing: Saarbrucken, Germany.

Errera Q, Ménard R. 2012. Technical note: spectral representation of spatial correlations in variational assimilation with grid point models and application to the Belgian Assimilation System for Chemical Observations (BASCOE). *Atmos. Chem. Phys.* **12**: 10015–10031.

Etherton BJ, Bishop CH. 2004. Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Weather Rev.* **132**: 1065–1080.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162, doi: 10.1029/94JC00572.

Evensen G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**: 343–367.

Evensen G. 2007. *Data Assimilation, The Ensemble Kalman Filter*. Springer-Verlag: Berlin.

Evensen G, van Leeuwen PJ. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**: 1852–1867.

Fairbairn D, Pring SR, Lorenc AC, Roulstone I. 2014. A comparison of 4D-Var with ensemble data assimilation methods. *Q. J. R. Meteorol. Soc.* **140**: 281–294.

Fertig EJ, Hunt BR, Ott E, Szunyough I. 2008. Assimilating non-local observations with a local ensemble Kalman Filter. *Tellus* **59A**: 719–730.

Fisher M. 1998. 'Development of a simplified Kalman filter', Technical Note 260. ECMWF: Reading, UK.

Fisher M. 2003. 'Background error covariance modelling'. In *Proceedings of the ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, 8–12 September 2003. ECMWF: Reading, UK.

Fisher M, Andersson A. 2001. 'Developments in 4D-Var and Kalman filtering', Technical Note 347. ECMWF: Reading, UK.

Fisher M, Courtier P. 1995. *Estimating the Covariance Matrices of Analysis and Forecast Error in Variational Data Assimilation*. Technical Memo 220. ECMWF: Reading, UK.

Fresnay S, Hally A, Garnaud C, Richard E, Lambert D. 2012. Heavy precipitation events in the Mediterranean: Sensitivity to cloud physics parameterisation uncertainties. *Nat. Hazards Earth Syst. Sci.* **12**: 2671–2688.

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757.

Gauthier P, Charette C, Fillion L, Koclas P, Laroche S. 1999. Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmosphere Ocean* **37**: 103–156.

Gauthier P, Tanguay M, Laroche S, Pellerin S, Morneau J. 2007. Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.* **135**: 2339–2354.

Ge G, Gao J, Xue M. 2012. Diagnostic pressure equation as a weak constraint in a storm-scale three-dimensional variational radar data assimilation system. *J. Atmos. Oceanic Technol.* **29**: 1075–1092.

Gebhardt C, Theis SE, Paulat M, Bouallgue ZB. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.* **100**: 168–177.

Goodliff M, Amezcua J, van Leeuwen PJ. 2015. Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing nonlinearity. *Tellus* **67A**: 26928, doi: 10.3402/tellusa.v67.26928.

Goodliff M, Amezcua J, van Leeuwen PJ. 2017. A weak-constraint 4DEnsembleVar. Part II: experiments with larger models. *Tellus* **69A**: 1271565, doi: 10.1080/16000870.2016.1271565.

Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR. 2011. Balance and ensemble Kalman filter localization techniques. *Mon. Weather Rev.* **139**: 511–522.

Gustafsson N. 2007. Discussion on '4D-Var or ensemble Kalman filter?' *Tellus* **59A**: 774–777.

Gustafsson N, Bojarova J. 2014. Four-dimensional ensemble variational (4D-En-Var) data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). *Nonlin. Proc. Geophys.* **21**: 745–762, doi: 10.5194/npg-21-745-2014.

Gustafsson N, Berre L, Hornquist S, Huang XY, Lindskog M, Navascues B, Mogensen KS, Thornsteinsson S. 2001. Three-dimensional variational data assimilation for a limited-area model. *Tellus* **53A**: 425–446.

Gustafsson N, Bojarova J, Vigne O. 2014. A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). *Nonlin. Proc. Geophys.* **21**: 303–323.

Haben S, Lawless AS, Nichols NK. 2011. Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus* **63A**: 782–792.

Hacker JP, Snyder C, Ha S-Y, Pocernich M. 2011. Linear and non-linear response to parameter variations in a mesoscale model. *Tellus* **63A**: 429–444.

Hamill TM. 2006. Ensemble-based atmospheric data assimilation. In *Predictability of Weather and Climate*, Palmer TN, Hagedorn R. (eds.): 124–156. Cambridge University Press: Cambridge, UK.

Hammill TM, Snyder C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Weather Rev.* **128**: 2905–2919.

Hamill TM, Whitaker JS. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Weather Rev.* **133**: 3132–3147.

Hamill TM, Whitaker JS, Synder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790.

Hamill TM, Whitaker JS, Fiorino M, Benjamin SG. 2011a. Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Weather Rev.* **139**: 668–688.

Hamill TM, Whitaker JS, Kleist DT, Fiorino M, Benjamin SG. 2011b. Predictions of 2010's tropical cyclones using the GFS and ensemble-based data assimilation methods. *Mon. Weather Rev.* **139**: 3243–3247.

Hayden CM, Purser RJ. 1995. Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteorol.* **34**: 3–13.

Hogan TF, Rosmond TE, Gelaro R. 1992. 'The NOGAPS forecast model: A technical description', Report AD-A247 216. Naval Research Laboratory: Monterey, CA. www.dtic.mil/dtic/tr/fulltext/u2/a247216.pdf (accessed 12 January 2017).

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**: 796–811.

Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.

Houtekamer PL, Mitchell HL. 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.* **131**: 3269–3289.

Houtekamer PL, Zhang F. 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **144**: 4489–4532.

Houtekamer PL, Lefaivre L, Derome J, Ritchie H, Mitchell HL. 1996. A system simulation approach to ensemble prediction. *Mon. Weather Rev.* **124**: 1225–1242.

Houtekamer PL, Pekkerin G, Buehner M, Charron M, Spacek L, Hansen B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Weather Rev.* **133**: 604–620.

Houtekamer PL, Deng X, Mitchell HL, Baek S-J, Gagnon N. 2014. Higher resolution in an operational ensemble Kalman filter. *Mon. Weather Rev.* **142**: 1143–1162.

Howes K. 2016. Accounting for model error in four-dimensional variational data assimilation. PhD thesis, Department of Mathematics, University of Reading: Reading, UK.

Hunt BR, Kalnay E, Kostelich EJ, Ott E, Patil DJ, Sauer T, Szunyogh I, Yorke JA, Zimin AV. 2004. Four-dimensional ensemble Kalman filtering. *Tellus* **56A**: 273–277.

Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: Operational, sequential and variational. *J. Meteorol. Soc. Jpn.* **75**: 181–189.

Ingleby NB. 2001. The statistical structure of forecast errors and its representation in the Met Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **127**: 209–231.

Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. *Ensemble of Data Assimilations at ECMWF*, Technical Memorandum 260. ECMWF: Reading, UK.

Jazwinski AH. 1970. *Stochastic Processes and Filtering Theory*, Academic Press: New York, NY.

Kalnay E. 2002. *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press: Cambridge, UK.

Kalnay E, Li H, Miyoshi T, Yang S-C, Ballabrera-Poy J. 2007. 4D-Var or ensemble Kalman filter? *Tellus* **59A**: 758–773.

Kepert JD. 2009. Covariance localization and balance in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 1157–1176.

Kleist DT, Ide K. 2015a. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Weather Rev.* **143**: 433–451.

Kleist DT, Ide K. 2015b. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Weather Rev.* **143**: 452–470.

Kleist DT, Parrish DF, Derber JC, Treadon R, Errico RM, Yang R. 2009. Improving incremental balance in the GSI 3DVAR analysis system. *Mon. Weather Rev.* **137**: 1046–1060.

Kuhl DD, Rosmond TE, Bishop CH, McLay J, Baker NL. 2013. Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Weather Rev.* **141**: 2740–2758.

Lawless AS. 2010. A note on the analysis error associated with 3D-FGAT. *Q. J. R. Meteorol. Soc.* **136**: 1094–1098, doi: 10.1002/qj.619.

Lawrence AR, Leutbecher M, Palmer TN. 2009. The characteristics of Hessian singular vectors using an advanced data assimilation scheme. *Q. J. R. Meteorol. Soc.* **135**: 1117–1132.

Le Dimet F-X, Talagrand O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus* **38A**: 97–110.

Lee M-S, Barker D, Huang W, Kuo Y-H. 2004. 'First guess at appropriate time (FGAT) with WRF 3D-Var'. In *Preprints for WRF/MM5 Users' Workshop, 22–25 June 2004. Boulder, CO. http://www.mmm.ucar.edu/mm5/workshop/ws04/Session5/Lee.Mi-Seon1.pdf (accessed 12 January 2017).*

van Leeuwen PJ. 1999. Comment on 'Data assimilation using an ensemble Kalman filter technique'. *Mon. Weather Rev.* **127**: 1374–1377.

van Leeuwen PJ. 2009. Particle filtering in geophysical systems. *Mon. Weather Rev.* **137**: 4089–4114, doi: 10.1175/2009MWR2835.1.

van Leeuwen PJ, Evensen G. 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* **124**: 2898–2913.

Leith CE. 1993. Numerical models of weather and climate. *Plasma Phys. Controlled Fusion* **35**: 919–927.

Liu DC, Nocedal J. 1989. On the limited memory BFGS method for large scale optimisation. *Math. Program.* **45**: 503–528.

Liu C, Xiao Q. 2013. An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with advanced research WRF using real data. *Mon. Weather Rev.* **141**: 2721–2739, doi: 10.1175/MWR-D-12-00130.1.

Liu C, Xue M. 2016. Relationships among four-dimensional hybrid ensemble-variational data assimilation algorithms with full and approximate ensemble localization. *Mon. Weather Rev.* **144**: 591–606, doi: 10.1175/MWR-D-15-0203.1.

Liu C, Xiao Q, Wang B. 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formation and preliminary test. *Mon. Weather Rev.* **136**: 3363–3373.

Liu C, Xiao Q, Wang B. 2009. An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with the advanced research WRF (ARW). *Mon. Weather Rev.* **137**: 1687–1704.

Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**: 3183–3203.

Lorenc AC. 2013. 'Recommended nomenclature for EnVar data assimilation methods'. In WGNE Blue Book Research Activities in Atmospheric and Oceanic Modelling, section 01: 7–8. WMO: Geneva, Switzerland. http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01_Lorenc_Andrew_EnVar_nomenclature.pdf (accessed 23 December 2016).

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **126**: 2991–3012.

Lorenc AC, Bowler NE, Clayton AM, Pring SR, Fairbairn D. 2015. Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weather Rev.* **143**: 212–229.

Lorenz EN. 1996. 'Predictability: A problem partly solved'. In Proceedings of the Seminar on Predictability, Vol. 1. ECMWF: Reading, UK.

Lorenz EN. 2005. Designing chaotic models. *J. Atmos. Sci.* **62**: 1574–1587.

Lorenz EN. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**: 130–148.

Ménétrier M, Auligné T. 2015. An overlooked issue of variational data assimilation. *Mon. Weather Rev.* **143**: 3925–3930, doi: 10.1175/MWR-D-14-00404.1.

Meng Z, Zhang F. 2007. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Weather Rev.* **135**: 1403–1423.

Meng Z, Zhang F. 2008a. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar in a real-data case study. *Mon. Weather Rev.* **136**: 522–540.

Meng Z, Zhang F. 2008b. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Weather Rev.* **136**: 3671–3682.

Meng Z, Zhang F. 2012. Ensemble-based data assimilation. In *Encyclopedia of Atmospheric Sciences* (2nd edn), North GR, Pyle JA, Zhang F. (eds.), Vol. 2: 241–247. Elsevier: Amsterdam, Netherlands.

Morss RE, Emanuel KA. 2002. Influence of added observations on analysis and forecast errors: Results from idealized systems. *Q. J. R. Meteorol. Soc.* **128**: 285–321.

Navon IM, Daescu DN, Liu Z. 2005. The impact of background error on incomplete observations for 4D-Var data assimilation with the FSU GSM. In *Computational Science–ICCS*, **2**: 837-844. Atlanta, GA.

NCEP Environmental Modeling Center. *The Global Forecast System (GFS) – Global Spectral Model (GSM)*. NOAA, NCEP: College Park, MD. http://www.emc.ncep.noaa.gov/GFS/doc.php (accessed 13 January 2017).

Oleson KW, Dai Y, Bonan G, Bosilovich M, Dickinson R, Dirmeyer P, Hoffman F, Houser P, Levis S, Niu G-Y, Thornton P, Vertenstein M, Yang Z-L, Zeng X. 2004. 'Technical description of the Community Land Model (CLM)', Technical Note NCAR/TN-461+STR. NCAR: Boulder, CO. http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLM_Tech_Note.pdf (accessed 13 January 2017).

Ott E, Hunt BR, Szunyough I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**: 415–428.

Paige CC, Saunders MA. 1975. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.* **12**: 617–629.

Palmer TN. 2012. Towards the probabilistic Earth system simulator: A vision for the future of climate and weather prediction. *Q. J. R. Meteorol. Soc.* **138**: 841–861.

Palmer TN, Williams P. 2010. *Stochastic Physics and Climate Modelling*. Cambridge University Press: Cambridge, UK.

Pan Y, Zhu K, Xue M, Wang X, Hu M, Benjamin SG, Weygandt SS, Whitaker JS. 2014. A GSI-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational rapid refresh model: Tests at a reduced resolution. *Mon. Weather Rev.* **142**: 3756–3780.

Park SK, Zupanski D. 2003. Four-dimensional variational data assimilation for mesoscale and storm scale applications. *Meteorol. Atmos. Phys.* **82**: 173–208.

Parrish DF, Derber JC. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **120**: 1747–1793.

Penny SG. 2014. The hybrid local ensemble transform Kalman filter. *Mon. Weather Rev.* **142**: 2139–2149, doi: 10.1175/mwr-d-13-00131.1.

Petrie RE, Bannister RN. 2011. A method for merging flow-dependent forecast error statistics from an ensemble with static statistics for use in high-resolution variational data assimilation. *Comput. Fluids* **46**: 387–391.

Pham DT. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.* **129**: 1194–1207.

Poterjoy J, Zhang F. 2014. Intercomparison and coupling of ensemble and four- dimensional variational data assimilation methods for the analysis and forecasting of hurricane *Karl* (2010). *Mon. Weather Rev.* **142**: 3347–3364.

Poterjoy J, Zhang F. 2015. Systematic comparison for four-dimensional data assimilation methods with and without the tangent linear model using hybrid background-error covariance: E4DVar versus 4DVar. *Mon. Weather Rev.* **143**: 1601–1621.

Rabier F. 2005. Overview of global data assimilation developments in numerical weather prediction centres. *Q. J. R. Meteorol. Soc.* **131**: 3215–3233.

Rabier F, Järvinen H, Klinker E, Mahfouf J-F, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **126**: 1143–1170.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.

Raynaud L, Berre L, Desroziers G. 2011. An extended specification of flow-dependent background-error variances in the Météo-France global 4D-Var system. *Q. J. R. Meteorol. Soc.* **137**: 607–619, doi: 10.1002/qj.795.

Raynaud L, Berre L, Desroziers G. 2012. Accounting for model error in the Météo-France ensemble data assimilation system. *Q. J. R. Meteorol. Soc.* **138**: 249–262.

Rhodin A. 2015. 'Multi Scale VarEnKF Localisation using Wavelets'. In *Presentation at International Symposium on Data Assimilation,*

RIKEN, February 2015. Kobe, Japan. http://www.data-assimilation.riken.jp/isda2015/program/pdf/8-4.pdf (accessed 13 January 2017).
Rodgers CD. 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing: Singapore.
Sasaki Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Weather Rev.* **98**: 875–883.
Schlatter TW. 2000. Variational assimilation of meteorological observations in the lower atmosphere: A tutorial on how it works. *J. Atmos. Sol. Terr. Phys.* **62**: 1057–1070.
Schwartz CS, Liu Z. 2014. Convection-permitting forecasts initialized with continuously cycling limited-area 3DVar, ensemble Kalman filter and 'hybrid' variational-ensemble data assimilation systems. *Mon. Weather Rev.* **142**: 716–737.
Schwartz CS, Liu Z, Lin H-C, McKeen SA. 2012. Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *J. Geophys. Res.* **117**: D13202, doi: 10.1029/2011JD017383.
Schwartz CS, Liu Z, Huang X-Y, Kuo Y-H. 2013. Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: Sensitivity to outer loops and vortex relocation. *Mon. Weather Rev.* **141**: 4350–4372.
Schwartz CS, Liu Z, Lin H-C, Cetola JD. 2014. Assimilating aerosol observations with a 'hybrid' variational-ensemble data assimilation system. *J. Geophys. Res. Atmos.* **119**: 4043–4069, doi: 10.1002/2013JD020937.
Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **131**: 3079–3102.
Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG. 2005. 'A description of the advanced research WRF version 2', Technical Note NCAR/TN-486+STR. NCAR: Boulder, CO.
Stensrud DJ, Bao J-W, Warner TT. 2000. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.* **128**: 2077–2107.
Stewart LM, Dance SL, Nichols NK, Eyre JR, Cameron J. 2014. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **140**: 1236–1244, doi: 10.1002/qj.2211.
Stiller O. 2009. Efficient moist physics schemes for data assimilation. II: Deep convection. *Q. J. R. Meteorol. Soc.* **135**: 721–738, doi: 10.1002/qj.362.
Stiller O, Ballard SP. 2009. Efficient moist physics schemes for data assimilation. I: Large-scale clouds and condensation. *Q. J. R. Meteorol. Soc.* **135**: 707–720, doi: 10.1002/qj.400.
Szunyogh I, Kostelich EJ, Gyarmati G, Patil DJ, Hunt BR, Kalnay E, Ott E, Yorke JA. 2005. Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus* **57A**: 528–545.
Szunyogh I, Kostelich EJ, Gyarmati G, Kalnay E, Hunt BR, Ott E, Satterfield E, Yorke JA. 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus* **60A**: 113–130.
Talagrand O, Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc.* **113**: 1311–1328.
Tarantola A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM: Philadelphia, PA.
Temperton C. 1988. Implicit normal mode initialization. *Mon. Weather Rev.* **116**: 1013–1031.
Thepaut J-N, Courtier P. 1991. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.* **117**: 1225–1254.
Tian X, Xie Z, Dai A. 2008. An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.* **113**: D21124, doi: 10.1029/2008JD010358.
Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Ensemble square root filters. *Mon. Weather Rev.* **131**: 1485–1490.
Trémolet Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**: 2483–2504.
Trémolet Y. 2007. Model-error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.* **133**: 1267–1280.
Tribbia JJ, Baumhefner DP. 2004. Scale interactions and atmospheric predictability: An updated perspective. *Mon. Weather Rev.* **132**: 703–713.
Undén P, Rontu L, Järvinen H, Lynch P, Calvo J, Cats G, Cuxart J, Eerola K, Fortelius C, Garcia-Moya JA, Jones C, Lenderlink G, McDonald A, McGrath R, Navascues B, Nielsen NW, Ødegaard V, Rodriguez E, Rummukainen M, Rõõm R, Sattler K, Sass BH, Savijärvi H, Schreur BW, Sigg R, The H, Tijm A. 2002. 'HIRLAM-5 scientific documentation (2002)', http://hirlam.org/index.php/hirlam-documentation (accessed 13 January 2017).
Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosazac M, Fisher M, Fuentes M, Hagemann S, Hólm EV, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KA, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012.
Vié B, Molinié G, Nuissier O, Vincendon B, Ducrocq V, Bouttier F, Richard E. 2012. Hydro-meteorological evaluation of a convection-permitting ensemble prediction system for Mediterranean heavy precipitating events. *Nat. Hazards Earth Syst. Sci.* **12**: 2631–2645.
Waller JA, Dance SL, Lawless AS, Nichols NK. 2014. Estimating correlated observation-error statistics using an ensemble transform Kalman filter. *Tellus* **66A**: 23294, doi: 10.3402/tellusa.v66.23294.
Wang X. 2010. Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. *Mon. Weather Rev.* **138**: 2990–2995.
Wang X. 2011. Application of the WRF hybrid ETKF-3DVAR data assimilation system for hurricane track forecasts. *Weather and Forecasting* **26**: 868–884.
Wang X, Lei T. 2014. GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon. Weather Rev.* **142**: 3303–3325.
Wang X, Hamill TM, Whitaker JS, Bishop CH. 2007a. A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Mon. Weather Rev.* **135**: 1055–1076.
Wang X, Snyder C, Hamill TM. 2007b. On the theoretical equivalence of differently proposed ensemble-3DVar hybrid analysis schemes. *Mon. Weather Rev.* **135**: 222–227.
Wang X, Barker DM, Snyder C, Hamill TM. 2008a. A hybrid ETKF-3D-Var data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **136**: 5116–5131.
Wang X, Barker DM, Snyder C, Hamill TM. 2008b. A hybrid ETKF-3D-Var data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Weather Rev.* **136**: 5132–5147.
Wang XG, Hamill TM, Whitaker JS, Bishop CH. 2009. A comparison of the hybrid and EnSRF analysis schemes in the presence of model errors due to unresolved scales. *Mon. Weather Rev.* **137**: 3219–3232.
Wang X, Parrish D, Kleist D, Whitaker J. 2013. GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP global forecast system: single-resolution experiments. *Mon. Weather Rev.* **141**: 4098–4117.
Wang H, Huang X-Y, Sun J, Xu D, Zhang M, Fan S, Zhong J. 2014. Inhomogeneous background error modeling for WRF-var using the NMC method. *J. Appl. Meterol. Climatol.* **53**: 2287–2309.
Weaver AT, Deltel C, Machu E, Ricci S, Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3605–3626.
Whitaker J, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**: 1913–1924.
Whitaker J, Compo GP, Wei X, Hamill TM. 2002. Reanalysis without radiosondes using ensemble data assimilation. *Mon. Weather Rev.* **132**: 1190–1200.
Whitaker J, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Weather Rev.* **436**: 463–482.
Wunsch C. 2012. *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications*. Cambridge University Press: Cambridge, UK.
Xu Q. 1996. Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.* **53**: 1123–1142.
Žagar N, Gustafsson N, Källén E. 2004. Dynamical response of equatorial waves in four-dimensional variational data assimilation. *Tellus* **56A**: 29–46.
Zangl G, Reinert D, Ripodas P, Baldauf M. 2015. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.* **141**: 563–579.
Zhang M, Zhang F. 2012. E4DVar: coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Weather Rev.* **140**: 587–600.
Zhang F, Snyder C, Sun J. 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **132**: 1238–1253.
Zhang F, Zhang M, Hansen JA. 2009. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.* **26**: 1–8.
Zhang M, Zhang F, Huang X-Y, Zhang X. 2011. Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. *Mon. Weather Rev.* **139**: 566–572.
Zhang F, Meng Z, Aksoy A. 2016. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Weather Rev.* **134**: 722–736.
Zhang F, Zhang M, Poterjoy J. 2013. E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Mon. Weather Rev.* **141**: 900–917.
Zupanski D. 1997. A general weak constraint applicable to operational 4D-Var data assimilation systems. *Mon. Weather Rev.* **125**: 2274–2291.
Zupanski M. 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Weather Rev.* **133**: 1710–1726.